

Supplementary Material for

“Panel Data Models with Interactive Fixed Effects and (Relatively) Small T ”

Ayden Higgins

September 22, 2026

Contents

A Additional Discussion	47
B Proofs of Main Results	52
C Proofs of Lemmas B.1 - B.3	72
D Results for Section 5.1	83
E Results for Section 5.3	99
F Higher Order Terms	120
G Additional Monte Carlo Material	125

Appendix A. Additional Material

A.1. Properties of $Q_{\mathcal{X}}$

The following properties are used throughout the paper, often without explicit reference. Assume $\mathcal{X}$ has full rank.

1. $Q_{\mathcal{X}}Q_{\mathcal{X}}^{\top} = P_{\mathcal{X}}$ and $P_{\mathcal{X}} = I_n$ if $TK \geq n$;
2. $Q_{\mathcal{X}}^{\top}Q_{\mathcal{X}} = I_{TK}$ if $TK \leq n$;
3. $P_{\mathcal{X}}\mathcal{X} = \mathcal{X}$;
4. $\text{tr}(P_{\mathcal{X}}) = \text{rank}(\mathcal{X})$;
5. $\|Q_{\mathcal{X}}\|_2 = \|P_{\mathcal{X}}\|_2 = 1$;
6. $\|Q_{\mathcal{X}}\|_F = \|P_{\mathcal{X}}\|_F = \sqrt{\text{rank}(P_{\mathcal{X}})}$;
7. $\text{rank}(Q_{\mathcal{X}}) = \text{rank}(P_{\mathcal{X}}) = \mathcal{O}(T)$.

Proof:

1. Notice $\mathbf{X}^\top \mathbf{X} = \mathbf{V} \mathbf{S}^2 \mathbf{V}^\top$ and therefore $(\mathbf{X}^\top \mathbf{X})^+ = \mathbf{V} \mathbf{S}^{-2} \mathbf{V}^\top$. Hence,

$$\mathbf{Q}_\mathbf{X} \mathbf{Q}_\mathbf{X}^\top = \mathbf{U} \mathbf{V}^\top \mathbf{V} \mathbf{U}^\top = \mathbf{U} \mathbf{S} \mathbf{V}^\top \mathbf{V} \mathbf{S}^{-2} \mathbf{V}^\top \mathbf{V} \mathbf{S} \mathbf{U}^\top = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top = \mathbf{P}_\mathbf{X}.$$

When $TK \geq n$, $\text{rank}(\mathbf{X}) = n$, whereby $\mathbf{U}$ is orthogonal, from which it follows $\mathbf{Q}_\mathbf{X} \mathbf{Q}_\mathbf{X}^\top = \mathbf{U} \mathbf{V}^\top \mathbf{V} \mathbf{U}^\top = \mathbf{U} \mathbf{U}^\top = \mathbf{I}_n$.

2. When $TK \leq n$, $\text{rank}(\mathbf{X}) = TK$, whereby $\mathbf{V}$ is orthogonal, from which it follows

$$\mathbf{Q}_\mathbf{X}^\top \mathbf{Q}_\mathbf{X} = \mathbf{V} \mathbf{V}^\top = \mathbf{V} \mathbf{V}^{-1} = \mathbf{I}_{TK}.$$

3. $\mathbf{P}_\mathbf{X} \mathbf{X} = \mathbf{U} \mathbf{U}^\top \mathbf{U} \mathbf{S} \mathbf{V}^\top = \mathbf{U} \mathbf{S} \mathbf{V}^\top = \mathbf{X}$.

4. $\text{tr}(\mathbf{P}_\mathbf{X}) = \text{tr}(\mathbf{U} \mathbf{U}^\top) = \text{tr}(\mathbf{U}^\top \mathbf{U}) = \text{tr}(\mathbf{I}_{\text{rank}(\mathbf{X})}) = \text{rank}(\mathbf{X})$.

5. $\|\mathbf{Q}_\mathbf{X}\|_2^2 = \mu_{\max}(\mathbf{Q}_\mathbf{X}^\top \mathbf{Q}_\mathbf{X}) = \mu_{\max}(\mathbf{Q}_\mathbf{X} \mathbf{Q}_\mathbf{X}^\top) = \mu_{\max}(\mathbf{U} \mathbf{U}^\top) = \mu_{\max}(\mathbf{P}_\mathbf{X}) = 1 = \|\mathbf{P}_\mathbf{X}\|_2^2$.

6. $\|\mathbf{Q}_\mathbf{X}\|_F^2 = \text{tr}(\mathbf{Q}_\mathbf{X} \mathbf{Q}_\mathbf{X}^\top) = \text{tr}(\mathbf{P}_\mathbf{X}) = \text{rank}(\mathbf{X})$.

7. $\text{rank}(\mathbf{Q}_\mathbf{X}) = \text{rank}(\mathbf{P}_\mathbf{X}) = \text{rank}(\mathbf{X}) = \min\{n, TK\} \leq TK = \mathcal{O}(T)$.

□

A.2. Equations (3.1) and (3.2)

To establish equation (3.1),

$$\begin{aligned} & \min_{\tilde{\mathbf{A}} \in \mathbb{R}^{TK \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \mu_{\min} \left(\frac{1}{nT} \tilde{\mathbf{X}}^\top (\mathbf{M}_\mathbf{F} \otimes \mathbf{M}_{\tilde{\mathbf{A}}}) \tilde{\mathbf{X}} \right) \\ &= \min_{\boldsymbol{\delta} \in \mathbb{R}^K: \|\boldsymbol{\delta}\|_2=1} \min_{\tilde{\mathbf{A}} \in \mathbb{R}^{TK \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \left(\frac{1}{nT} (\tilde{\mathbf{X}} \boldsymbol{\delta})^\top (\mathbf{M}_\mathbf{F} \otimes \mathbf{M}_{\tilde{\mathbf{A}}}) \tilde{\mathbf{X}} \boldsymbol{\delta} \right) \\ &= \min_{\boldsymbol{\delta} \in \mathbb{R}^K: \|\boldsymbol{\delta}\|_2=1} \min_{\tilde{\mathbf{A}} \in \mathbb{R}^{TK \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \frac{1}{nT} \text{tr} \left((\tilde{\mathbf{X}} \cdot \boldsymbol{\delta})^\top \mathbf{M}_{\tilde{\mathbf{A}}} (\tilde{\mathbf{X}} \cdot \boldsymbol{\delta}) \mathbf{M}_\mathbf{F} \right) \\ &= \min_{\boldsymbol{\delta} \in \mathbb{R}^K: \|\boldsymbol{\delta}\|_2=1} \sum_{t=T_{\min}}^T \mu_t \left(\frac{1}{nT} (\tilde{\mathbf{X}} \cdot \boldsymbol{\delta})^\top (\tilde{\mathbf{X}} \cdot \boldsymbol{\delta}) \right), \end{aligned}$$

where the last line follows by Lemma A.1 in Moon and Weidner (2017). For (3.2), using the previous display,

$$\begin{aligned} & \min_{\boldsymbol{\delta} \in \mathbb{R}^K: \|\boldsymbol{\delta}\|_2=1} \sum_{t=T_{\min}}^T \mu_t \left(\frac{1}{nT} (\tilde{\mathbf{X}} \cdot \boldsymbol{\delta})^\top (\tilde{\mathbf{X}} \cdot \boldsymbol{\delta}) \right) \\ &= \min_{\tilde{\mathbf{A}} \in \mathbb{R}^{TK \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \mu_{\min} \left(\frac{1}{nT} \mathbf{X}^\top (\mathbf{M}_\mathbf{F} \otimes \mathbf{Q}_\mathbf{X} \mathbf{M}_{\tilde{\mathbf{A}}} \mathbf{Q}_\mathbf{X}^\top) \mathbf{X} \right) \\ &= \min_{\mathbf{A} \in \mathbb{R}^{n \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \mu_{\min} \left(\frac{1}{nT} \mathbf{X}^\top (\mathbf{M}_\mathbf{F} \otimes (\mathbf{P}_\mathbf{X} - \mathbf{P}_\mathbf{X} \mathbf{A} (\mathbf{A}^\top \mathbf{P}_\mathbf{X} \mathbf{A})^+ \mathbf{A}^\top \mathbf{P}_\mathbf{X}) \mathbf{X} \right) \\ &= \min_{\mathbf{A} \in \mathbb{R}^{n \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \mu_{\min} \left(\frac{1}{nT} \mathbf{X}^\top (\mathbf{M}_\mathbf{F} \otimes \mathbf{M}_{\mathbf{P}_\mathbf{X} \mathbf{A}}) \mathbf{X} \right) \\ &\geq \min_{\mathbf{A} \in \mathbb{R}^{n \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \mu_{\min} \left(\frac{1}{nT} \mathbf{X}^\top (\mathbf{M}_\mathbf{F} \otimes \mathbf{M}_\mathbf{A}) \mathbf{X} \right). \end{aligned}$$

A.3. Equation (3.4)

Equation (3.4) is demonstrated to hold under the following assumption.

Assumption AX.

- (i) The elements of $\mathbf{\Lambda}_1$ have uniformly bounded second moments and $n^{-1} \tilde{\mathbf{\Lambda}}_1^\top \tilde{\mathbf{\Lambda}}_1 \xrightarrow{p} \mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_1}$ as $n \rightarrow \infty$, with $T \geq R_1 + 1$ fixed or $T \rightarrow \infty$, where the eigenvalues of $\mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_1}$ are bounded away from zero and from above by a constant.
- (ii) $\boldsymbol{\lambda}_{2,i} \sim \text{iid}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\lambda}_2})$ and is independent of the covariates and $\mathbf{\Lambda}_1$, with uniformly bounded fourth moment, and the minimum eigenvalue of $\mathbf{\Sigma}_{\boldsymbol{\lambda}_2}$ bounded away from zero.

First,

$$\frac{1}{n} \tilde{\mathbf{\Lambda}}_1^\top \tilde{\mathbf{\Lambda}}_1 \xrightarrow{p} \mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_1}, \quad (\text{A.1})$$

as $n \rightarrow \infty$, with T fixed or $T \rightarrow \infty$ and $T/n \rightarrow \gamma \in [0, \infty)$ by Assumption AX(i). Second,

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n^2} \left(\tilde{\mathbf{\Lambda}}_1^\top \tilde{\mathbf{\Lambda}}_2 \right)_{r_1, r_2}^2 \right] &= \frac{1}{n^2} \mathbb{E} \left[\mathbb{E} \left[\boldsymbol{\lambda}_{1, r_1}^\top \mathbf{P} \mathbf{x} \boldsymbol{\lambda}_{2, r_2} \boldsymbol{\lambda}_{2, r_2}^\top \mathbf{P} \mathbf{x} \boldsymbol{\lambda}_{1, r_1} \mid \mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{\Lambda}_1 \right] \right] \\ &= \frac{1}{n^2} (\mathbf{\Sigma}_{\boldsymbol{\lambda}_2})_{r_2, r_2} \mathbb{E} \left[\boldsymbol{\lambda}_{1, r_1}^\top \mathbf{P} \mathbf{x} \boldsymbol{\lambda}_{1, r_1} \right] \\ &\leq \frac{1}{n^2} (\mathbf{\Sigma}_{\boldsymbol{\lambda}_2})_{r_2, r_2} \mathbb{E} [\|\boldsymbol{\lambda}_{1, r_1}\|_2^2] = \mathcal{O} \left(\frac{1}{n} \right), \end{aligned}$$

using Assumptions AX(i) and AX(ii), from which it follows that

$$\frac{1}{n} \tilde{\mathbf{\Lambda}}_1^\top \tilde{\mathbf{\Lambda}}_2 \xrightarrow{p} \mathbf{0}, \quad (\text{A.2})$$

as $n \rightarrow \infty$, with T fixed or $T \rightarrow \infty$ and $T/n \rightarrow \gamma \in [0, \infty)$. Finally, notice that

$$\begin{aligned} \mathbb{E} \left[\mathbb{E} \left[\frac{1}{n} \left(\tilde{\mathbf{\Lambda}}_2^\top \tilde{\mathbf{\Lambda}}_2 \right)_{r_1, r_2} \mid \mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{\Lambda}_1 \right] \right] &= \frac{1}{n} \mathbb{E} \left[\mathbb{E} \left[\boldsymbol{\lambda}_{2, r_1}^\top \mathbf{P} \mathbf{x} \boldsymbol{\lambda}_{2, r_2} \mid \mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{\Lambda}_1 \right] \right] \\ &= \frac{1}{n} (\mathbf{\Sigma}_{\boldsymbol{\lambda}_2})_{r_1, r_2} \mathbb{E} [\text{tr}(\mathbf{P} \mathbf{x})] \\ &= \frac{1}{n} \min\{n, TK\} (\mathbf{\Sigma}_{\boldsymbol{\lambda}_2})_{r_1, r_2}, \end{aligned}$$

by Assumption **AX(ii)** and since $\text{tr}(\mathbf{P}\mathbf{x}) = \min\{n, TK\}$. Using this result,

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{1}{n} \left(\tilde{\mathbf{A}}_2^\top \tilde{\mathbf{A}}_2 \right)_{r_1, r_2} - \frac{1}{n} \min\{n, TK\} (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2} \right)^2 \right] \\
&= \mathbb{E} \left[\left(\frac{1}{n} \left(\tilde{\mathbf{A}}_2^\top \tilde{\mathbf{A}}_2 \right)_{r_1, r_2} \right)^2 \right] - \frac{2}{n} \min\{n, TK\} (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2} \mathbb{E} \left[\frac{1}{n} \left(\tilde{\mathbf{A}}_2^\top \tilde{\mathbf{A}}_2 \right)_{r_1, r_2} \right] \\
&\quad + \frac{1}{n^2} \min\{n, TK\}^2 (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \\
&= \frac{1}{n^2} \mathbb{E} \left[\mathbb{E} \left[\lambda_{2, r_1}^\top \mathbf{P}\mathbf{x} \lambda_{2, r_2} \lambda_{2, r_1}^\top \mathbf{P}\mathbf{x} \lambda_{2, r_2} \mid \mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{\Lambda}_1 \right] \right] - \frac{1}{n^2} \min\{n, TK\}^2 (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \\
&= \frac{1}{n^2} \min\{n, TK\}^2 (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \\
&\quad + \frac{1}{n^2} \min\{n, TK\} \left((\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_1} (\boldsymbol{\Sigma}_{\lambda_2})_{r_2, r_2} + (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \right) \\
&\quad + \frac{1}{n^2} \left(\mathbb{E} [\lambda_{2, i, r_1}^2 \lambda_{2, i, r_2}^2] - (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_1} (\boldsymbol{\Sigma}_{\lambda_2})_{r_2, r_2} - 2(\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \right) \mathbb{E} \left[\sum_i (\mathbf{P}\mathbf{x})_{ii}^2 \right] \\
&\quad - \frac{1}{n^2} \min\{n, TK\}^2 (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \\
&= \frac{1}{n^2} \min\{n, TK\} \left((\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_1} (\boldsymbol{\Sigma}_{\lambda_2})_{r_2, r_2} + (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \right) \\
&\quad + \frac{1}{n^2} \left(\mathbb{E} [\lambda_{2, i, r_1}^2 \lambda_{2, i, r_2}^2] - (\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_1} (\boldsymbol{\Sigma}_{\lambda_2})_{r_2, r_2} - 2(\boldsymbol{\Sigma}_{\lambda_2})_{r_1, r_2}^2 \right) \mathbb{E} \left[\sum_i (\mathbf{P}\mathbf{x})_{ii}^2 \right] \\
&= \mathcal{O} \left(\frac{1}{n} \right),
\end{aligned}$$

which follows by Assumption **AX(ii)** and since

$$\sum_i (\mathbf{P}\mathbf{x})_{ii}^2 \leq \sum_i (\mathbf{P}\mathbf{x})_{ii} = \text{tr}(\mathbf{P}\mathbf{x}) = \min\{n, TK\} \leq n.$$

Thus,

$$\frac{1}{n} \tilde{\mathbf{A}}_2^\top \tilde{\mathbf{A}}_2 - \min \left\{ 1, \frac{TK}{n} \right\} \boldsymbol{\Sigma}_{\lambda_2} \xrightarrow{p} \mathbf{0},$$

as $n \rightarrow \infty$, with T fixed or $T \rightarrow \infty$ and $T/n \rightarrow \gamma \in [0, \infty)$, from which it follows that

$$\frac{1}{n} \tilde{\mathbf{A}}_2^\top \tilde{\mathbf{A}}_2 \xrightarrow{p} \min\{1, \gamma K\} \times \boldsymbol{\Sigma}_{\lambda_2}. \quad \text{³⁹ (A.3)}$$

Combining (A.1), (A.2), and (A.3) yields the result.

A.4. Footnote 27

Assume $\text{rank}(\tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top) = R_0$ and $1 \leq R_0 \leq T$. Suppose it is possible to decompose $\tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top = \tilde{\mathbf{\Lambda}}_* \mathbf{F}_*^\top$ such that

$$\mathbf{F}_* = \begin{pmatrix} \mathbf{I}_{R_0} \\ \boldsymbol{\Phi} \end{pmatrix},$$

³⁹ Recall the convention that $\gamma := \lim_{n \rightarrow \infty} T/n = 0$ when T is fixed.

where Φ is $(T - R_0) \times R_0$. Setting

$$\mathbf{v}_0 := \begin{pmatrix} \Phi^\top \\ -\mathbf{I}_{T-R_0} \end{pmatrix},$$

one has

$$\mathbf{v}_0^\top \mathbf{F}_* \tilde{\Lambda}_*^\top = \begin{pmatrix} \Phi & -\mathbf{I}_{T-R_0} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{R_0} \\ \Phi \end{pmatrix} \tilde{\Lambda}_*^\top = (\Phi - \Phi) \tilde{\Lambda}_*^\top = \mathbf{0}.$$

Now, instead, suppose

$$\mathbf{v}_0 = \begin{pmatrix} \Phi^\top \\ -\mathbf{I}_{T-R_0} \end{pmatrix}$$

and $\mathbf{v}_0^\top \mathbf{F}_0 \tilde{\Lambda}_0^\top = \mathbf{0}$. Then

$$\mathbf{v}_0^\top \mathbf{F}_0 \tilde{\Lambda}_0^\top = \begin{pmatrix} \Phi & -\mathbf{I}_{T-R_0} \end{pmatrix} \begin{pmatrix} \mathbf{F}_{0,1} \\ \mathbf{F}_{0,2} \end{pmatrix} \tilde{\Lambda}_0^\top = \Phi \mathbf{F}_{0,1} \tilde{\Lambda}_0^\top - \mathbf{F}_{0,2} \tilde{\Lambda}_0^\top = \mathbf{0}.$$

Hence, $\Phi \mathbf{F}_{0,1} \tilde{\Lambda}_0^\top = \mathbf{F}_{0,2} \tilde{\Lambda}_0^\top$, whereby

$$\tilde{\Lambda}_0 \mathbf{F}_0^\top = \begin{pmatrix} \tilde{\Lambda}_0 \mathbf{F}_{0,1}^\top & \tilde{\Lambda}_0 \mathbf{F}_{0,2}^\top \end{pmatrix} = \begin{pmatrix} \tilde{\Lambda}_0 \mathbf{F}_{0,1}^\top & \tilde{\Lambda}_0 \mathbf{F}_{0,1}^\top \Phi^\top \end{pmatrix} = \tilde{\Lambda}_0 \mathbf{F}_{0,1}^\top \begin{pmatrix} \mathbf{I}_{R_0} & \Phi^\top \end{pmatrix} = \tilde{\Lambda}_* \mathbf{F}_*^\top.$$

A.5. Assumptions **IF(iii)** and **IF(iv)**

The section provides an example of error dependence which satisfies Assumptions **IF(iii)** and **IF(iv)**. For simplicity it is assumed that ε_{it} is independent of $\mathcal{D}_{nT}$. Suppose $\varepsilon_{it} = \phi \varepsilon_{it-1} + u_{it}$ with $|\phi| < 1$, and u_{it} independent over i and t with $\mathbb{E}[u_{it}] = 0$, $\mathbb{E}[u_{it}^2] = \sigma_u^2$ and $\mathbb{E}[u_{it}^{16}]$ uniformly bounded. Moreover, assume $\varepsilon_{i0} = 0$. The extension to heterogeneous ϕ , heteroskedastic u_{it} , higher order autoregressive processes, and nonzero initial conditions comes at the expense of more cumbersome notation. Write

$$\mathbf{v}^\top \boldsymbol{\varepsilon}_i = \mathbf{v}^\top \mathbf{S}^{-\top} \mathbf{u}_i =: \mathbf{a}^\top \mathbf{u}_i,$$

where $\mathbf{S} := \mathbf{I}_T - \phi \mathbf{W}$ and $\mathbf{W}$ is a $T \times T$ shift matrix which consists of zeros everywhere, except those elements directly above the main diagonal, which take on a value of 1. Using the Rosenthal inequality 9.7.c in [Lin \(2010\)](#),

$$\mathbb{E} \left[|\mathbf{v}^\top \boldsymbol{\varepsilon}_i|^{16} \right] \lesssim \left(\sum_{t=1}^T \mathbb{E}[a_t^2 u_{it}^2] \right)^8 + \sum_{t=1}^T \mathbb{E} \left[|a_t u_{it}|^{16} \right]. \quad (\text{A.4})$$

Now,

$$\sum_{t=1}^T \mathbb{E}[a_t^2 u_{it}^2] \lesssim \left(\sum_{t=1}^T a_t^2 \right) \leq \|\mathbf{S}^{-1}\|_2^2,$$

and

$$\sum_{t=1}^T \mathbb{E} [|a_t u_{it}|^{16}] \lesssim \left(\sum_{t=1}^T a_t^{16} \right) \leq \|\mathbf{S}^{-1}\|_2^{16}.$$

Using $\|\mathbf{S}^{-1}\|_2 \leq c$, which follows by the same arguments as in the proof of Lemma E.1(ii), the right-hand side of (A.4) is uniformly bounded which verifies IF(iii). For IF(iv), notice that

$$\mathbb{E}[\varepsilon_{it_1} \varepsilon_{it_2}] = (1 - \phi^{2 \min\{t_1, t_2\}}) \frac{\sigma_u^2}{1 - \phi^2} \phi^{|t_1 - t_2|} =: b(t_1, t_2) \phi^{|t_1 - t_2|},$$

whereby

$$\mathbb{E}[\varepsilon_{i_1 t_1} \varepsilon_{i_2 t_2}] = \begin{cases} b(t_1, t_2) \phi^{|t_1 - t_2|} & \text{for } i_1 = i_2 \\ 0 & \text{otherwise,} \end{cases}$$

and thereby $\Delta_{\mathcal{D}} := \Gamma_{b_T}(\Sigma_{\mathcal{D}}) - \Sigma_{\mathcal{D}}$ is such that

$$\delta_{\mathcal{D}, i_1 i_2 t_1 t_2} = \begin{cases} -b(t_1, t_2) \phi^{|t_1 - t_2|} & \text{for } i_1 = i_2 \text{ and } |t_1 - t_2| \geq b_T \\ 0 & \text{otherwise.} \end{cases}$$

Using this,

$$\begin{aligned} \|\Delta_{\mathcal{D}}\|_2 &\leq \|\Delta_{\mathcal{D}}\|_1 \leq \max_{1 \leq t_1 \leq T} \sum_{t_2=1}^T 1\{|t_1 - t_2| \geq b_T\} b(t_1, t_2) |\phi|^{|t_1 - t_2|} \\ &\leq \max_{1 \leq t_1 \leq T} \sum_{t_2 < t_1}^T 1\{t_1 - t_2 \geq b_T\} b(t_1, t_2) |\phi|^{t_1 - t_2} \\ &\quad + \max_{1 \leq t_1 \leq T} \sum_{t_2 > t_1}^T 1\{t_2 - t_1 \geq b_T\} b(t_1, t_2) |\phi|^{t_2 - t_1} \\ &\lesssim |\phi|^{b_T} \rightarrow 0, \end{aligned}$$

as $b_T \rightarrow \infty$. This verifies IF(iv).

Appendix B. Proofs of Main Results

Notation

Throughout this appendix and subsequent appendices, some additional notation is employed. Let $\sigma_{\mathcal{A}, it}^2 := \mathbb{E}[\varepsilon_{it}^2 | \mathcal{A}_{nT}]$ and $\xi_{\mathcal{A}, it}^j := \mathbb{E}[\varepsilon_{it}^j | \mathcal{A}_{nT}]$ for $j > 2$. Moreover, $a \lesssim b$ is used to indicate there exists a constant c such that $a \leq cb$. For an $n \times 1$ vector $\mathbf{a}$ with elements a_i , $\|\mathbf{a}\|_1 := \sum_{i=1}^n |a_i|$ and $\|\mathbf{a}\|_\infty := \max_{1 \leq i \leq n} |a_i|$. Let $\mathbf{A}$ be an $n \times m$ matrix with elements A_{ij} . The following matrix norms are those induced by their vector counterparts: $\|\mathbf{A}\|_1 := \max_{1 \leq j \leq m} \sum_{i=1}^n |A_{ij}|$, which is the maximum absolute column sum of $\mathbf{A}$, and $\|\mathbf{A}\|_\infty := \max_{1 \leq i \leq n} \sum_{j=1}^m |A_{ij}|$, which is the maximum absolute row sum of $\mathbf{A}$.

Lemmas

Lemma B.1. Under Assumption **MD**,

- (i) $\|\mathbf{X}_k\|_2, \|\tilde{\mathbf{X}}_k\|_2, \|\mathbf{X}_k\|_F, \|\tilde{\mathbf{X}}_k\|_F = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K$;
- (ii) $\|\mathbf{F}_0\|_2, \|\mathbf{F}_0\|_F = \mathcal{O}_p(\sqrt{T})$ and $\|\mathbf{\Lambda}_0\|_2, \|\mathbf{\Lambda}_0\|_F, \|\tilde{\mathbf{\Lambda}}_0\|_2, \|\tilde{\mathbf{\Lambda}}_0\|_F = \mathcal{O}_p(\sqrt{n})$.

Lemma B.2. Under Assumption **MD**, and Assumption **EC** or **ED**,

- (i) $(nT)^{-\frac{1}{2}}\|\tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p(n^{-\frac{1}{4}})$;
- (ii) $\text{tr}(\tilde{\mathbf{X}}_k^\top \tilde{\boldsymbol{\varepsilon}}) = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K$.

Lemma B.3. Under Assumptions **MD**, **ED**, **AE**, and **AD**,

- (i) $\mathbf{b}^{(1)} = -\boldsymbol{\psi}^{(1)} + \mathcal{O}_p(n^{-\frac{1}{4}})$;
- (ii) $\mathbf{b}^{(2)} = \mathcal{O}_p(n^{-\frac{1}{2}})$;
- (iii) $\mathbf{b}^{(3)} = -\boldsymbol{\psi}^{(2)} + \mathcal{O}_p(n^{-\frac{1}{2}})$.

Proof of Proposition 1.

Substituting the true DGP (2.1) into the expression for $\mathcal{Q}(\boldsymbol{\beta}, \tilde{\mathbf{\Lambda}}, \mathbf{F})$ (given in equation (2.3)) yields

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\beta}, \tilde{\mathbf{\Lambda}}, \mathbf{F}) &= \frac{1}{nT} \text{tr} \left(\left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top - \tilde{\mathbf{\Lambda}} \mathbf{F}^\top + \tilde{\boldsymbol{\varepsilon}} \right)^\top \right. \\ &\quad \left. \times \left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top - \tilde{\mathbf{\Lambda}} \mathbf{F}^\top + \tilde{\boldsymbol{\varepsilon}} \right) \right). \end{aligned} \quad (\text{B.1})$$

Now,

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\beta}) &= \min_{\tilde{\mathbf{\Lambda}} \in \mathbb{R}^{TK \times Re}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \frac{1}{nT} \text{tr} \left(\left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top - \tilde{\mathbf{\Lambda}} \mathbf{F}^\top + \tilde{\boldsymbol{\varepsilon}} \right)^\top \right. \\ &\quad \left. \times \left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top - \tilde{\mathbf{\Lambda}} \mathbf{F}^\top + \tilde{\boldsymbol{\varepsilon}} \right) \right) \\ &\geq \min_{\bar{\mathbf{F}} \in \mathbb{R}^{T \times (Re + R_0)}} \frac{1}{nT} \text{tr} \left(\left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \tilde{\boldsymbol{\varepsilon}} \right) \mathbf{M}_{\bar{\mathbf{F}}} \left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \tilde{\boldsymbol{\varepsilon}} \right)^\top \right) \\ &= \min_{\bar{\mathbf{F}} \in \mathbb{R}^{T \times (Re + R_0)}} \left\{ \frac{1}{nT} \text{tr} \left(\left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \right) \mathbf{M}_{\bar{\mathbf{F}}} \left(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \right)^\top \right) \right. \\ &\quad \left. + \frac{2}{nT} \sum_{k=1}^K (\beta_{0,k} - \beta_k) \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{M}_{\bar{\mathbf{F}}} \tilde{\mathbf{X}}_k^\top) + \frac{1}{nT} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{M}_{\bar{\mathbf{F}}} \tilde{\boldsymbol{\varepsilon}}^\top) \right\}, \end{aligned} \quad (\text{B.2})$$

which follows because the minimum can be no less than if one were able to minimise with respect to $\tilde{\mathbf{\Lambda}}$ and $\bar{\mathbf{F}}$, where $\tilde{\mathbf{\Lambda}} \bar{\mathbf{F}}^\top := \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top - \tilde{\mathbf{\Lambda}} \mathbf{F}^\top$ which can have rank of at most $R_e + R_0$ under Assumption **CS(i)**.

For the last two terms in (B.2),

$$\begin{aligned} \frac{2}{nT} \sum_{k=1}^K (\beta_{0,k} - \beta_k) \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{M}_{\bar{\mathbf{F}}} \tilde{\mathbf{X}}_k^\top) &= \frac{2}{nT} \sum_{k=1}^K (\beta_{0,k} - \beta_k) \text{tr}(\tilde{\boldsymbol{\varepsilon}} \tilde{\mathbf{X}}_k^\top) - \frac{2}{nT} \sum_{k=1}^K (\beta_{0,k} - \beta_k) \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{P}_{\bar{\mathbf{F}}} \tilde{\mathbf{X}}_k^\top) \\ &= \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right) + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \mathcal{O}_p \left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}} \right), \end{aligned} \quad (\text{B.3})$$

and

$$\begin{aligned}\frac{1}{nT}\text{tr}(\tilde{\boldsymbol{\varepsilon}}\mathbf{M}_{\bar{\mathbf{F}}}\tilde{\boldsymbol{\varepsilon}}^\top) &= \frac{1}{nT}\text{tr}(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^\top) - \frac{1}{nT}\text{tr}(\tilde{\boldsymbol{\varepsilon}}\mathbf{P}_{\bar{\mathbf{F}}}\tilde{\boldsymbol{\varepsilon}}^\top), \\ &= \frac{1}{nT}\text{tr}(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^\top) + \mathcal{O}_p\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2^2}{nT}\right),\end{aligned}\tag{B.4}$$

uniformly over $\bar{\mathbf{F}} \in \mathbb{R}^{T \times (R_e + R_0)}$ using Lemmas B.1(i), B.1(ii), and B.2(ii), as well as the inequality $|\text{tr}(\mathbf{A})| \leq \text{rank}(\mathbf{A})\|\mathbf{A}\|_2$. Applying (B.3) and (B.4) to (B.2),

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\beta}) &\geq \min_{\bar{\mathbf{F}} \in \mathbb{R}^{T \times (R_e + R_0)}} \left\{ \frac{1}{nT}\text{tr}\left((\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}))\mathbf{M}_{\bar{\mathbf{F}}}(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \boldsymbol{\beta}))^\top\right) \right\} \\ &\quad + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \mathcal{O}_p\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}}\right) + \mathcal{O}_p\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2^2}{nT}\right) + \frac{1}{nT}\text{tr}(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^\top) \\ &\geq c\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2 + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \mathcal{O}_p\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}}\right) + \mathcal{O}_p\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2^2}{nT}\right) + \frac{1}{nT}\text{tr}(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^\top)\end{aligned}\tag{B.5}$$

w.p.a.1, where the last inequality follows by Assumption CS(ii). Now, note that

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\beta}_0) &= \min_{\tilde{\mathbf{A}} \in \mathbb{R}^{TK \times R_e}, \mathbf{F} \in \mathbb{R}^{T \times R_0}} \frac{1}{nT}\text{tr}((\tilde{\mathbf{A}}_0\mathbf{F}_0^\top - \tilde{\mathbf{A}}\mathbf{F}^\top + \tilde{\boldsymbol{\varepsilon}})(\tilde{\mathbf{A}}_0\mathbf{F}_0^\top - \tilde{\mathbf{A}}\mathbf{F}^\top + \tilde{\boldsymbol{\varepsilon}})^\top) \\ &\leq \mathcal{Q}(\boldsymbol{\beta}_0, \tilde{\mathbf{A}}_0, \mathbf{F}_0) = \frac{1}{nT}\text{tr}(\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}}).\end{aligned}\tag{B.6}$$

Since $\mathcal{Q}(\hat{\boldsymbol{\beta}}) \leq \mathcal{Q}(\boldsymbol{\beta}_0)$ by definition, combining (B.5) and (B.6) gives

$$c\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2^2 + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \mathcal{O}_p\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}}\right) + \mathcal{O}_p\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2^2}{nT}\right) \leq 0,$$

w.p.a.1, from which the result is obtained by applying Lemma B.2(i). $\square$

Proof of Proposition 2.

Begin by defining the following

$$\begin{aligned}\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \boldsymbol{\beta} &= \tilde{\mathbf{A}}_0\mathbf{F}_0^\top + \tilde{\boldsymbol{\varepsilon}} + \sum_{k=1}^K (\beta_{0,k} - \beta_k)\tilde{\mathbf{X}}_k \\ &=: \tilde{\mathbf{A}}_0\mathbf{F}_0^\top + \sum_{k=0}^K e_k \tilde{\mathbf{X}}_k,\end{aligned}\tag{B.7}$$

with $e_0 := (nT)^{-\frac{1}{2}}\|\tilde{\boldsymbol{\varepsilon}}\|_2$, $e_k := (\beta_{0,k} - \beta_k)$ for $k = 1, \dots, K$, and $\tilde{\mathbf{X}}_0 := (nT)^{\frac{1}{2}}\|\tilde{\boldsymbol{\varepsilon}}\|_2^{-1}\tilde{\boldsymbol{\varepsilon}}$. Now, substituting (B.7) into the concentrated objective function (2.4), $\mathcal{Q}(\boldsymbol{\beta})$ is equal to

$$\begin{aligned}&\frac{1}{nT} \sum_{t=R_0+1}^T \mu_t \left(\mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top + \sum_{k=0}^K e_k \tilde{\mathbf{X}}_k^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top + \sum_{k=0}^K e_k \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{X}}_k + \sum_{k_1=0}^K \sum_{k_2=0}^K e_{k_1} e_{k_2} \tilde{\mathbf{X}}_{k_1}^\top \tilde{\mathbf{X}}_{k_2} \right) \\ &=: \sum_{t=R_0+1}^T \mu_t \left(\mathbf{T}^{(0)} + \mathbf{T}^{(1)}(\mathbf{e}) + \mathbf{T}^{(2)}(\mathbf{e}) \right).\end{aligned}$$

One may consider $\mathbf{T}(\mathbf{e}) := \mathbf{T}^{(0)} + \mathbf{T}^{(1)}(\mathbf{e}) + \mathbf{T}^{(2)}(\mathbf{e})$ to be a perturbation of the linear operator $\mathbf{T}^{(0)}$.

When the perturbation $\mathbf{e}$ is small, it may be possible to express the eigenvalues of $\mathbf{T}(\mathbf{e})$ as a power series

in $\mathbf{e}$. Define

$$\begin{aligned}\mu_{\max}(\tilde{\mathbf{A}}_0, \mathbf{F}_0) &:= \sqrt{\frac{1}{nT} \mu_1 \left(\mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top \right)} \\ \mu_{\min}(\tilde{\mathbf{A}}_0, \mathbf{F}_0) &:= \sqrt{\frac{1}{nT} \mu_{R_0} \left(\mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top \right)},\end{aligned}$$

and

$$r_0(\tilde{\mathbf{A}}_0, \mathbf{F}_0) := \left(\frac{4\mu_{\max}(\tilde{\mathbf{A}}_0, \mathbf{F}_0)}{\mu_{\min}(\tilde{\mathbf{A}}_0, \mathbf{F}_0)} + \frac{1}{2\mu_{\max}(\tilde{\mathbf{A}}_0, \mathbf{F}_0)} \right)^{-1}.$$

An expansion of $\mathcal{Q}(\beta)$ can be obtained under the condition

$$\frac{1}{\sqrt{nT}} \sum_{k=0}^K |e_k| \|\tilde{\mathbf{X}}_k\|_2 \leq r_0(\tilde{\mathbf{A}}_0, \mathbf{F}_0).^{40} \quad (\text{B.8})$$

Intuitively, this condition ensures the perturbation of $\mathbf{T}^{(0)}$ is sufficiently small. Now, $r_0(\tilde{\mathbf{A}}_0, \mathbf{F}_0)$ converges in probability to a positive constant by Assumptions [AE\(ii\)](#) and [AE\(iii\)](#). For $k = 1, \dots, K$,

$$\frac{1}{\sqrt{nT}} \|\tilde{\mathbf{X}}_k\|_2 = \mathcal{O}_p(1), \quad (\text{B.9})$$

by Lemma [B.1\(i\)](#). For $k = 0$, $(nT)^{-\frac{1}{2}} \|\tilde{\mathbf{X}}_0\|_2 = \|\tilde{\boldsymbol{\varepsilon}}\|_2^{-1} \|\tilde{\boldsymbol{\varepsilon}}\|_2 = 1$. Thus, $(nT)^{-\frac{1}{2}} \|\tilde{\mathbf{X}}_k\|_2 = \mathcal{O}_p(1)$ for $k = 0, \dots, K$ and so the satisfaction of [\(B.8\)](#) rests upon the extent to which $e_0, \dots, e_K$ are small. By Lemma [B.2\(i\)](#), $e_0 = (nT)^{-\frac{1}{2}} \|\tilde{\boldsymbol{\varepsilon}}\|_2$ is $\mathcal{O}_p(1)$. Also $e_k = \beta_{0,k} - \beta_k = \mathcal{O}_p(1)$ is assumed in the statement of the proposition. Thus, Lemma S.1 in [Moon and Weidner \(2017\)](#) is applicable with the result that $\mathcal{Q}(\beta)$ can be expanded in the following power series in $e_0, \dots, e_K$:

$$\mathcal{Q}(\beta) = \frac{1}{nT} \sum_{g=2}^{\infty} \sum_{k_1=0}^K \sum_{k_2=0}^K \dots \sum_{k_g=0}^K e_{k_1} e_{k_2} \dots e_{k_g} \mathcal{Q}^{(g)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \dots, \tilde{\mathbf{X}}_{k_g}), \quad (\text{B.10})$$

where

$$\mathcal{Q}^{(g)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \dots, \tilde{\mathbf{X}}_{k_g}) = \frac{1}{g!} \left(\tilde{\mathcal{Q}}^{(g)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \dots, \tilde{\mathbf{X}}_{k_g}) + \text{all permutations of } k_1, \dots, k_g \right),$$

with

$$\tilde{\mathcal{Q}}^{(g)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \dots, \tilde{\mathbf{X}}_{k_g}) := \sum_{p=1}^g (-1)^{p+1} \sum_{\substack{v_1 + \dots + v_p = g \\ m_1 + \dots + m_{p+1} = p-1 \\ 1 \leq v_j \leq 2 \\ m_j \geq 0}} \text{tr}(\mathbf{S}^{(m_1)} \mathcal{T}_{k_1 \dots}^{(v_1)} \mathbf{S}^{(m_2)} \dots \mathcal{T}_{\dots k_g}^{(v_p)} \mathbf{S}^{(m_{p+1})}),$$

⁴⁰See S.6.1 in [Moon and Weidner \(2017\)](#) for details.

and $\mathbf{S}^{(0)} := -\mathbf{M}_{\mathbf{F}_0}$, $\mathbf{S}^{(m)} := (\mathbf{F}_0(\mathbf{F}_0^\top \mathbf{F}_0)^{-1}(\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1}(\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top)^m$ for $m \geq 1$, $\mathcal{T}_k^{(1)} := \mathbf{F}_0 \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{X}}_k + \tilde{\mathbf{X}}_k^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top$, and $\mathcal{T}_{k_1, k_2}^{(2)} := \tilde{\mathbf{X}}_{k_1}^\top \tilde{\mathbf{X}}_{k_2}$. Moreover, under the stronger condition

$$\frac{1}{\sqrt{nT}} \sum_{k=0}^K |e_k| \|\tilde{\mathbf{X}}_k\|_2 \leq \frac{\mu_{\min}^2(\tilde{\mathbf{\Lambda}}_0, \mathbf{F}_0)}{16\mu_{\max}(\tilde{\mathbf{\Lambda}}_0, \mathbf{F}_0)}, \quad (\text{B.11})$$

this series can be truncated at order $G \geq 2$, with the following bound on the remainder

$$\left| \mathcal{Q}(\beta) - \frac{1}{nT} \sum_{g=2}^G \sum_{k_1=0}^K \sum_{k_2=0}^K \dots \sum_{k_g=0}^K e_{k_1} e_{k_2} \dots e_{k_g} \mathcal{Q}^{(g)}(\tilde{\mathbf{\Lambda}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \dots, \tilde{\mathbf{X}}_{k_g}) \right| \leq \frac{R_0(G+1)\varrho^{G+1} \mu_{\min}^2(\tilde{\mathbf{\Lambda}}_0, \mathbf{F}_0)}{2(1-\varrho)^2},$$

where,

$$\varrho := \frac{16\mu_{\max}(\tilde{\mathbf{\Lambda}}_0, \mathbf{F}_0)}{\mu_{\min}^2(\tilde{\mathbf{\Lambda}}_0, \mathbf{F}_0)} \left(\frac{1}{\sqrt{nT}} \sum_{k=0}^K |e_k| \|\tilde{\mathbf{X}}_k\|_2 \right) \leq 1. \quad (\text{B.12})$$

Since the right-hand side of (B.11) will converge in probability to a positive constant under Assumptions **AE(ii)** and **AE(iii)**, this condition will be satisfied w.p.a.1, and the above bound on the remainder becomes applicable. Hence, using the expansion (B.10) taken to order $G = 5$ one obtains

$$\mathcal{Q}(\beta) = \mathcal{Q}(\beta_0) - \frac{2}{\sqrt{nT}} (\beta - \beta_0)^\top \mathbf{d} + (\beta - \beta_0)^\top \mathbf{D} (\beta - \beta_0) + r(\beta), \quad (\text{B.13})$$

where $\mathbf{d} := \mathbf{c} + \mathbf{b}^{(1)} + \mathbf{b}^{(2)} + \mathbf{b}^{(3)}$, and elements of these vectors and matrices are

$$\begin{aligned} D_{k_1 k_2} &:= \frac{1}{nT} \text{tr}(\tilde{\mathbf{X}}_{k_1} \mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{X}}_{k_2}^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0}) \\ c_k &:= \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\mathbf{X}}_k \mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{\epsilon}}^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0}) \\ b_k^{(1)} &:= -\frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{\epsilon}}^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \tilde{\mathbf{X}}_k \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\epsilon}} \right) \\ b_k^{(2)} &:= -\frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{X}}_k^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \tilde{\mathbf{\epsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\epsilon}} \right) \\ b_k^{(3)} &:= -\frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{\epsilon}}^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \tilde{\mathbf{\epsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{X}}_k \right), \end{aligned}$$

for $k, k_1, k_2 = 1, \dots, K$, and

$$\begin{aligned}
r(\beta) := & \frac{1}{nT} \sum_{k_1=1}^K \sum_{k_2=1}^K e_{k_1} e_{k_2} e_0 \mathcal{Q}^{(3)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_0) \\
& + \frac{1}{nT} \sum_{k_1=1}^K \sum_{k_2=1}^K e_0 e_{k_1} e_{k_2} \mathcal{Q}^{(3)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}) \\
& + \frac{1}{nT} \sum_{k_1=1}^K \sum_{k_2=1}^K e_{k_1} e_0 e_{k_2} \mathcal{Q}^{(3)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_2}) \\
& + \frac{1}{nT} \sum_{k_1=1}^K \sum_{k_2=1}^K \sum_{k_3=1}^K e_{k_1} e_{k_2} e_{k_3} \mathcal{Q}^{(3)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}) \\
& + \frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}) \\
& + \frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K \sum_{k_5=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} e_{k_5} \mathcal{Q}^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}, \tilde{\mathbf{X}}_{k_5}) \\
& - \frac{1}{nT} e_0 e_0 e_0 e_0 \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
& - \frac{1}{nT} e_0 e_0 e_0 e_0 \mathcal{Q}^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
& + \mathcal{O}_p \left(\left(\sum_{k=1}^K |\beta_{0,k} - \beta_k| \frac{\|\tilde{\mathbf{X}}_k\|_2}{\sqrt{nT}} + \frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}} \right)^6 \right) - \mathcal{O}_p \left(\left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}} \right)^6 \right), \tag{B.14}
\end{aligned}$$

where the last term in (B.15) is not just some term of that order, but exactly the term of that order contained in $((nT)^{-\frac{1}{2}} \sum_{k=1}^K |\beta_{0,k} - \beta_k| \|\tilde{\mathbf{X}}_k\|_2 + (nT)^{-\frac{1}{2}} \|\tilde{\boldsymbol{\varepsilon}}\|_2)^6$. This term is already accounted for in $\mathcal{Q}(\beta_0)$. Now, consider the remainder term $r(\beta)$ which is expressed in (B.14). For the first term in (B.14),

$$\begin{aligned}
& \frac{1}{nT} \left| \sum_{k_1=1}^K \sum_{k_2=1}^K e_{k_1} e_{k_2} e_0 \mathcal{Q}^{(3)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_0) \right| \\
& \lesssim \frac{1}{n^2 T^2} \left(\sum_{k=1}^K |\beta_{0,k} - \beta_k| \|\tilde{\mathbf{X}}_k\|_2 \right)^2 \left\| \tilde{\mathbf{A}}_0 \right\|_2 \left\| \left(\frac{1}{n} \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \|\mathbf{F}_0\|_2 \|\tilde{\boldsymbol{\varepsilon}}\|_2 \\
& = \mathcal{O}_p \left(\|\beta_0 - \beta\|_2^2 \frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}} \right),
\end{aligned}$$

using Lemmas B.1(i) and B.1(ii), as well as Assumptions AE(ii) and AE(iii). The second and third terms in (B.14) are both of the same order. Using Lemmas B.1(i) and B.1(ii), as well as Assumptions

AE(ii) and **AE(iii)**, the fourth term is $\mathcal{O}_p(\|\beta_0 - \beta\|_2^3)$. Thus,

$$\begin{aligned}
r(\beta) &= \mathcal{O}_p \left(\|\beta_0 - \beta\|_2^2 \frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right) + \mathcal{O}_p(\|\beta_0 - \beta\|_2^3) \\
&+ \frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} \mathcal{Q}^{(4)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}) \\
&+ \frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K \sum_{k_5=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} e_{k_5} \mathcal{Q}^{(5)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}, \tilde{\mathbf{X}}_{k_5}) \\
&- \frac{1}{nT} e_0 e_0 e_0 e_0 \mathcal{Q}^{(4)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
&- \frac{1}{nT} e_0 e_0 e_0 e_0 e_0 \mathcal{Q}^{(5)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
&+ \mathcal{O}_p \left(\left(\sum_{k=1}^K |\beta_{0,k} - \beta_k| \frac{\|\tilde{\mathbf{X}}_k\|_2}{\sqrt{nT}} + \frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right)^6 \right) - \mathcal{O}_p \left(\left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right)^6 \right). \tag{B.15}
\end{aligned}$$

The second and third terms on the right-hand side of (B.15) are studied in Appendix F, where it is established that

$$\begin{aligned}
&\frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} \mathcal{Q}^{(4)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}) \\
&+ \frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K \sum_{k_5=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} e_{k_5} \mathcal{Q}^{(5)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}, \tilde{\mathbf{X}}_{k_5}) \\
&= \frac{1}{nT} e_0 e_0 e_0 e_0 \mathcal{Q}^{(4)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) + \frac{1}{nT} e_0 e_0 e_0 e_0 e_0 \mathcal{Q}^{(5)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
&+ \mathcal{O}_p(\|\beta - \beta_0\|_2^5) + \mathcal{O}_p \left(\|\beta - \beta_0\|_2^4 \frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right) + \mathcal{O}_p(\|\beta - \beta_0\|_2^4) + \mathcal{O}_p \left(\|\beta - \beta_0\|_2^3 \left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right)^2 \right) \\
&+ \mathcal{O}_p \left(\|\beta - \beta_0\|_2^3 \frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right) + \mathcal{O}_p \left(\|\beta - \beta_0\|_2^2 \left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right)^3 \right) + \mathcal{O}_p \left(\|\beta - \beta_0\|_2^2 \left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right)^2 \right) \\
&+ \mathcal{O}_p \left(\frac{\|\beta - \beta_0\|_2}{\sqrt{nT}} \right).
\end{aligned}$$

Combining the above with (B.13) and (B.15)

$$\begin{aligned}
\mathcal{Q}(\beta) &= \mathcal{Q}(\beta_0) - \frac{2}{\sqrt{nT}} (\beta - \beta_0)^\top \mathbf{d} + (\beta - \beta_0)^\top \mathbf{D} (\beta - \beta_0) \\
&+ \frac{1}{nT} \left(nT \|\beta_0 - \beta\|_2^2 \times \mathcal{O}_p(1) + \sqrt{nT} \|\beta_0 - \beta\|_2 \times \mathcal{O}_p(1) \right),
\end{aligned}$$

using Lemma B.2(i). □

Proof of Theorem 1.

Notice that by Assumption AD(ii), the matrix $\mathbf{D}$ is invertible and $\|\mathbf{D}^{-1}\|_2 \leq c$ w.p.a.1. Define $\mathbf{w} := (nT)^{-\frac{1}{2}} \mathbf{D}^{-1} \mathbf{d}$. The stochastic order of this quantity is examined. First, by assumption, $\|\mathbf{c}\|_2 = \mathcal{O}_p(1)$. By Lemmas B.3(i), B.3(ii), and B.3(iii),

$$\mathbf{b}^{(1)}, \mathbf{b}^{(3)} = \mathcal{O}_p \left(\sqrt{\frac{T}{n}} \right),$$

and $\mathbf{b}^{(2)} = \mathcal{O}_p(n^{-\frac{1}{2}})$. Hence $\|\mathbf{w}\|_2 = \mathcal{O}_p((nT)^{-\frac{1}{2}}) = o_p(1)$. Now, notice that Proposition 1 goes through replacing Assumption EC with ED (Lemmas B.2(i) and B.2(ii) hold under either assumption). Thus, applying Proposition 2,

$$\begin{aligned}\mathcal{Q}(\hat{\beta}) &= \mathcal{Q}(\beta_0) + (\hat{\beta} - \beta_0 - \mathbf{w})^\top \mathbf{D}(\hat{\beta} - \beta_0 - \mathbf{w}) - \mathbf{w}^\top \mathbf{D} \mathbf{w} + \mathbf{r}(\hat{\beta}) \\ \mathcal{Q}(\beta_0 + \mathbf{w}) &= \mathcal{Q}(\beta_0) - \mathbf{w}^\top \mathbf{D} \mathbf{w} + \mathbf{r}(\beta_0 + \mathbf{w}).\end{aligned}$$

Since $\mathcal{Q}(\hat{\beta}) \leq \mathcal{Q}(\beta_0 + \mathbf{w})$,

$$(\hat{\beta} - \beta_0 - \mathbf{w})^\top \mathbf{D}(\hat{\beta} - \beta_0 - \mathbf{w}) \leq \mathbf{r}(\beta_0 + \mathbf{w}) - \mathbf{r}(\hat{\beta}). \quad (\text{B.16})$$

Given $\|\mathbf{w}\|_2 = o_p(1)$, the order of the remainder term established in Proposition 2 is valid. Thus (B.16) becomes

$$\begin{aligned}(\hat{\beta} - \beta_0 - \mathbf{w})^\top \mathbf{D}(\hat{\beta} - \beta_0 - \mathbf{w}) \\ \leq o_p\left(\frac{1}{nT}\right) \left(\mathcal{O}_p((1 + \sqrt{nT}\|\mathbf{w}\|_2)^2) + \mathcal{O}_p((1 + \sqrt{nT}\|\beta_0 - \hat{\beta}\|_2)^2) \right),\end{aligned} \quad (\text{B.17})$$

Multiplying (B.17) by nT , using $\|\hat{\beta} - \beta_0 - \mathbf{w}\|_2^2 \mu_{\min}(\mathbf{D}) \leq (\hat{\beta} - \beta_0 - \mathbf{w})^\top \mathbf{D}(\hat{\beta} - \beta_0 - \mathbf{w})$, Assumption CS(ii), and the inequality $\sqrt{(x+y)} \leq \sqrt{x} + \sqrt{y}$ gives

$$\sqrt{nT}\|\hat{\beta} - \beta_0 - \mathbf{w}\|_2 \leq o_p(1 + \sqrt{nT}\|\mathbf{w}\|_2) + o_p(1 + \sqrt{nT}\|\beta_0 - \hat{\beta}\|_2) = o_p(1),$$

and, in turn,

$$\sqrt{nT}(\hat{\beta} - \beta_0) = \mathbf{D}^{-1}(\mathbf{c} + \mathbf{b}^{(1)} + \mathbf{b}^{(3)}) + o_p(1), \quad (\text{B.18})$$

since $\mathbf{b}^{(2)} = \mathcal{O}_p(n^{-\frac{1}{2}})$. To obtain the final result, let $\mathbf{v}$ be a $K \times 1$ vector with $\|\mathbf{v}\|_2 = 1$. Consider,

$$\begin{aligned}\mathbf{v}^\top \mathbb{V}^{-\frac{1}{2}} \mathbf{c} &= \frac{1}{\sqrt{nT}} \mathbf{v}^\top \mathbb{V}^{-\frac{1}{2}} \mathbf{X}^\top (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathbf{X}} \mathbf{M}_{\Lambda_0} \mathbf{Q}_{\mathbf{X}}^\top) \text{vec}(\boldsymbol{\varepsilon}) \\ &=: \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T a_{it} \varepsilon_{it} \\ &= \sum_{i=1}^n \left(\frac{1}{\sqrt{nT}} \sum_{t=1}^T a_{it} \varepsilon_{it} \right) \\ &=: \sum_{i=1}^n b_i.\end{aligned} \quad (\text{B.19})$$

Let $\mathcal{F}_i := \mathcal{D}_{nT} \vee \sigma(\varepsilon_1, \dots, \varepsilon_i)$ for $i \geq 1$ and $\mathcal{F}_0 := \mathcal{D}_{nT}$. Then $\{\mathcal{F}_i : 1 \leq i \leq n\}$ is such that b_i is $\mathcal{F}_i$ -measurable, $\mathcal{F}_{i-1} \subseteq \mathcal{F}_i$, and $\mathbb{E}[b_i | \mathcal{F}_{i-1}] = 0$ by Assumption ED. Moreover, let $\mathbf{X}_k^* := \mathbf{M}_{\mathbf{P}_{\mathbf{X}} \Lambda_0} \mathbf{X}_k \mathbf{M}_{\mathbf{F}_0}$. Then

$$\mathbb{E}[|b_i|] \leq \frac{1}{\sqrt{nT}} \mathbb{E} \left[\|\mathbf{a}_i\|_2 \times \left| \mathbf{d}_i^\top \boldsymbol{\varepsilon}_i \right| \right] \lesssim \frac{1}{\sqrt{nT}} \mathbb{E} [\|\mathbf{X}_i^*\|_2] \leq c,$$

with $\mathbf{d}_i := \mathbf{a}_i / \|\mathbf{a}_i\|_2$ for $\|\mathbf{a}_i\|_2 > 0$, and $\mathbf{d}_i := \mathbf{0}$ otherwise, using Assumptions ED and AD(ii). Thus $\{b_i, \mathcal{F}_i\}$ forms a martingale difference array. The conditions for the application of Theorem 3.2 in Hall

and Heyde (1980) are now verified. First,

$$\mathbb{E} \left[\max_{1 \leq i \leq n} |b_i| \right] \leq \left(\mathbb{E} \left[\sum_{i=1}^n b_i^4 \right] \right)^{\frac{1}{4}} \leq n^{\frac{1}{4}} \max_{1 \leq i \leq n} (\mathbb{E}[b_i^4])^{\frac{1}{4}} = \mathcal{O}(n^{-\frac{1}{4}}),$$

since

$$\mathbb{E}[b_i^4] = \frac{1}{n^2 T^2} \mathbb{E} \left[\left(\|\mathbf{a}_i\|_2 \times \left| \mathbf{d}_i^\top \boldsymbol{\varepsilon}_i \right| \right)^4 \right] \lesssim \frac{1}{n^2 T^2} \mathbb{E} [\|\mathbf{X}_i^*\|_2^4] \lesssim \frac{1}{n^2}, \quad (\text{B.20})$$

using Assumptions ED and AD(ii), from which it follows that $\max_{1 \leq i \leq n} |b_i| = \mathcal{O}_p(1)$. Second,

$$\begin{aligned} \sum_{i=1}^n b_i^2 - 1 &= \frac{1}{nT} \sum_{i=1}^n \mathbf{v}^\top \mathbb{V}^{-\frac{1}{2}} (\mathbf{X}_i^*)^\top \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \mathbf{X}_i^* \mathbb{V}^{-\frac{1}{2}} \mathbf{v} - \mathbf{v}^\top \mathbb{V}^{-\frac{1}{2}} \mathbb{V} \mathbb{V}^{-\frac{1}{2}} \mathbf{v} \\ &= \mathbf{v}^\top \mathbb{V}^{-\frac{1}{2}} \left(\frac{1}{nT} \sum_{i=1}^n (\mathbf{X}_i^*)^\top \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \mathbf{X}_i^* - \mathbb{V} \right) \mathbb{V}^{-\frac{1}{2}} \mathbf{v} \\ &= \mathbf{v}^\top \mathbb{V}^{-\frac{1}{2}} \left(\frac{1}{nT} \sum_{i=1}^n (\mathbf{X}_i^*)^\top \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \mathbf{X}_i^* - \mathbf{V} \right) \mathbb{V}^{-\frac{1}{2}} \mathbf{v} + \mathcal{O}_p(1) \\ &= \mathbf{v}^\top \mathbb{V}^{-\frac{1}{2}} \left(\frac{1}{nT} \sum_{i=1}^n (\mathbf{X}_i^*)^\top (\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top - \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \mathbf{X}_i^* \right) \mathbb{V}^{-\frac{1}{2}} \mathbf{v} + \mathcal{O}_p(1), \end{aligned} \quad (\text{B.21})$$

using Assumption AD(ii). Consider element k_1, k_2 of the term in brackets and, for simplicity, set $K = 1$.

$$\begin{aligned} &\frac{1}{n^2 T^2} \mathbb{E} \left[\left(\sum_{i=1}^n (\mathbf{x}_i^*)^\top (\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top - \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \mathbf{x}_i^* \right)^2 \right] \\ &= \frac{1}{n^2 T^2} \mathbb{E} \left[\sum_{i=1}^n \left((\mathbf{x}_i^*)^\top (\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top - \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \mathbf{x}_i^* \right)^2 \right] \\ &= \frac{1}{n^2 T^2} \mathbb{E} \left[\sum_{i=1}^n (\boldsymbol{\varepsilon}_i^\top \mathbf{x}_i^*)^4 \right] - \frac{1}{n^2 T^2} \mathbb{E} \left[\sum_{i=1}^n \left((\mathbf{x}_i^*)^\top \boldsymbol{\Sigma}_{\mathcal{D}, ii} \mathbf{x}_i^* \right)^2 \right] \\ &\lesssim \frac{1}{n^2 T^2} \mathbb{E} \left[\sum_{i=1}^n \|\mathbf{x}_i^*\|_2^4 \right] = \mathcal{O} \left(\frac{1}{n} \right), \end{aligned} \quad (\text{B.22})$$

using Assumptions ED and AD(i). Hence, under Assumption AD(ii), the right-hand side of (B.21) is $\mathcal{O}_p(1)$. Thirdly,

$$\mathbb{E} \left[\max_{1 \leq i \leq n} b_i^2 \right] \leq \left(\mathbb{E} \left[\max_{1 \leq i \leq n} |b_i|^4 \right] \right)^{\frac{1}{2}} \leq \left(\mathbb{E} \left[\sum_{i=1}^n |b_i|^4 \right] \right)^{\frac{1}{2}} \leq n^{\frac{1}{2}} \max_{1 \leq i \leq n} (\mathbb{E}[b_i^4])^{\frac{1}{2}} \leq c,$$

follows using (B.20). Therefore, by Theorem 3.2 in Hall and Heyde (1980) and with application of Assumption AD(ii),

$$\mathbf{D}^{-1} \mathbf{c} \xrightarrow{d} \mathcal{N}(0, \mathbb{D}^{-1} \mathbb{V} \mathbb{D}^{-1}).$$

Finally, applying Lemmas B.3(i) and B.3(iii) completes the result. $\square$

Proof of Proposition 3.

The result can be obtained as a special case of Proposition 4, by noticing that $(M_{F_0} \otimes M_{\Lambda_0})^+ = (M_{F_0}^+ \otimes M_{\Lambda_0}^+) = (M_{F_0} \otimes M_{\Lambda_0})$. Similarly for $(M_{F_0} \otimes M_{\tilde{\Lambda}_0})$.

□

Proof of Proposition 4.

First rewrite

$$\begin{aligned}
& \tilde{\mathcal{X}}^\top \left((M_{F_0} \otimes M_{\tilde{\Lambda}_0}) \tilde{\Sigma} (M_{F_0} \otimes M_{\tilde{\Lambda}_0}) \right)^+ \tilde{\mathcal{X}} \\
&= \mathcal{X}^\top \left((M_{F_0} \otimes Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top) \Sigma (M_{F_0} \otimes Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top) \right)^+ \mathcal{X} \\
&= \mathcal{X}^\top (M_{F_0} \otimes Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top) \left((M_{F_0} \otimes Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top) \Sigma (M_{F_0} \otimes Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top) \right)^+ \\
&\quad \times (M_{F_0} \otimes Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top) \mathcal{X} \\
&=: \mathcal{X}^\top \Sigma^{-\frac{1}{2}} \tilde{A} (\tilde{A}^\top \tilde{A})^+ \tilde{A}^\top \Sigma^{-\frac{1}{2}} \mathcal{X},
\end{aligned}$$

and

$$\begin{aligned}
& \mathcal{X}^\top ((M_{F_0} \otimes M_{\Lambda_0}) \Sigma (M_{F_0} \otimes M_{\Lambda_0}))^+ \mathcal{X} \\
&= \mathcal{X}^\top (M_{F_0} \otimes M_{\Lambda_0}) ((M_{F_0} \otimes M_{\Lambda_0}) \Sigma (M_{F_0} \otimes M_{\Lambda_0}))^+ (M_{F_0} \otimes M_{\Lambda_0}) \mathcal{X} \\
&=: \mathcal{X}^\top \Sigma^{-\frac{1}{2}} A (A^\top A)^+ A^\top \Sigma^{-\frac{1}{2}} \mathcal{X}.
\end{aligned}$$

Thus, the result follows provided

$$\mathcal{X}^\top \Sigma^{-\frac{1}{2}} \left(A (A^\top A)^+ A^\top - \tilde{A} (\tilde{A}^\top \tilde{A})^+ \tilde{A}^\top \right) \Sigma^{-\frac{1}{2}} \mathcal{X} \succeq \mathbf{0}, \quad (\text{B.23})$$

which holds if middle matrix on the left is positive semidefinite. This is now established. First,

$$\begin{aligned}
M_{\Lambda_0} Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top &= (I_n - P_{\Lambda_0}) (P_{\mathcal{X}} - P_{\mathcal{X}} \Lambda_0 (\Lambda_0^\top P_{\mathcal{X}} \Lambda_0)^+ \Lambda_0^\top P_{\mathcal{X}}) \\
&= P_{\mathcal{X}} - P_{\mathcal{X}} \Lambda_0 (\Lambda_0^\top P_{\mathcal{X}} \Lambda_0)^+ \Lambda_0^\top P_{\mathcal{X}} - P_{\Lambda_0} P_{\mathcal{X}} \\
&\quad + P_{\Lambda_0} P_{\mathcal{X}} \Lambda_0 (\Lambda_0^\top P_{\mathcal{X}} \Lambda_0)^+ \Lambda_0^\top P_{\mathcal{X}} \\
&= P_{\mathcal{X}} - P_{\mathcal{X}} \Lambda_0 (\Lambda_0^\top P_{\mathcal{X}} \Lambda_0)^+ \Lambda_0^\top P_{\mathcal{X}} - P_{\Lambda_0} P_{\mathcal{X}} + P_{\Lambda_0} P_{\mathcal{X}} \\
&= Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top.
\end{aligned} \quad (\text{B.24})$$

Similarly, it can be shown that

$$Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top M_{\Lambda_0} = Q_{\mathcal{X}} M_{\tilde{\Lambda}_0} Q_{\mathcal{X}}^\top. \quad (\text{B.25})$$

Using (B.24) and (B.25),

$$\begin{aligned}
& \mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top \\
&= \Sigma^{\frac{1}{2}}(\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0})((\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0})\Sigma(\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0}))^+ (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0})\Sigma \\
&\quad \times (\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \left((\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma(\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \right)^+ \\
&\quad \times (\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma^{\frac{1}{2}} \\
&= \Sigma^{\frac{1}{2}}(\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0})((\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0})\Sigma(\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0}))^+ (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0})\Sigma \\
&\quad \times (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0}\mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \left((\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma(\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \right)^+ \\
&\quad \times (\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma^{\frac{1}{2}} \\
&= \Sigma^{\frac{1}{2}}(\mathbf{M}_{F_0} \otimes \mathbf{M}_{\Lambda_0}\mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \left((\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma(\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \right)^+ \\
&\quad \times (\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma^{\frac{1}{2}} \\
&= \Sigma^{\frac{1}{2}}(\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \left((\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma(\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top) \right)^+ \\
&\quad \times (\mathbf{M}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}}\mathbf{M}_{\tilde{\Lambda}_0}\mathbf{Q}_{\mathcal{X}}^\top)\Sigma^{\frac{1}{2}} \\
&= \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top.
\end{aligned}$$

In the same way one can show $\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top \mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top = \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top$. It therefore follows that

$$\begin{aligned}
& (\mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top - \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top)(\mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top - \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top) \\
&= \mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top - \mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top - \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top \mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top \\
&\quad + \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top \\
&= \mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top - \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top,
\end{aligned}$$

which establishes $\mathbf{A}(\mathbf{A}^\top \mathbf{A})^+ \mathbf{A}^\top - \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^+ \tilde{\mathbf{A}}^\top$ is idempotent, which coupled with the fact that it is symmetric, establishes it to be an orthogonal projector and therefore must be positive semidefinite. This proves (B.23). $\square$

Proof of Proposition 5.

With $\mathbf{W} = (\mathbf{I}_T \otimes nT(\mathcal{X}^\top \mathcal{X})^{-1})$, the minimiser of $\mathcal{Q}_{\text{RS}}(\boldsymbol{\beta}, \tilde{\mathbf{\Lambda}}, \mathbf{F})$ with respect to $\tilde{\mathbf{\Lambda}}$ is given by

$$\hat{\tilde{\mathbf{\Lambda}}}_{\text{RS}}(\boldsymbol{\beta}, \mathbf{F}(\boldsymbol{\Phi})) = (\mathcal{X}^\top \mathcal{X})^{-\frac{1}{2}} \mathcal{X}^\top (\mathbf{Y} - \mathbf{X} \cdot \boldsymbol{\beta}) \mathbf{F}(\boldsymbol{\Phi}) (\mathbf{F}^\top(\boldsymbol{\Phi}) \mathbf{F}(\boldsymbol{\Phi}))^{-1}.$$

Substituting this into the moment condition (4.3),

$$\begin{aligned}
\varphi_{\text{RS}}(\boldsymbol{\beta}, \hat{\tilde{\mathbf{\Lambda}}}_{\text{RS}}(\boldsymbol{\beta}, \mathbf{F}(\boldsymbol{\Phi})), \mathbf{F}(\boldsymbol{\Phi})) &= \text{vec}(\mathcal{X}^\top (\mathbf{Y} - \mathbf{X} \cdot \boldsymbol{\beta}) \mathbf{M}_{\mathbf{F}(\boldsymbol{\Phi})}) \\
&=: \varphi_{\text{RS}}(\boldsymbol{\beta}, \mathbf{F}(\boldsymbol{\Phi})),
\end{aligned}$$

and thereby

$$\min_{\tilde{\mathbf{A}} \in \Theta_{\tilde{\mathbf{A}}}} \mathcal{Q}_{\text{RS}}(\boldsymbol{\beta}, \tilde{\mathbf{A}}, \mathbf{F}(\Phi)) = \boldsymbol{\varphi}_{\text{RS}}^\top(\boldsymbol{\beta}, \mathbf{F}(\Phi)) \mathbf{W} \boldsymbol{\varphi}_{\text{RS}}(\boldsymbol{\beta}, \mathbf{F}(\Phi)).$$

Now,

$$\sqrt{nT} \mathbf{W}^{\frac{1}{2}} \boldsymbol{\varphi}_{\text{RS}}(\boldsymbol{\beta}, \mathbf{F}(\Phi)) = \text{vec}(\mathbf{Q}_{\mathcal{X}}^\top (\mathbf{Y} - \mathbf{X} \cdot \boldsymbol{\beta}) \mathbf{M}_{\mathbf{F}(\Phi)}).$$

with which

$$\begin{aligned} \sqrt{nT} \mathbf{W}^{\frac{1}{2}} \mathbb{E}[\nabla \boldsymbol{\varphi}_{\text{RS}}(\boldsymbol{\beta}_0, \mathbf{F}_*) | \mathcal{D}_{nT}] &= \begin{pmatrix} -(\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathcal{X}})^\top \boldsymbol{\chi} & -(\mathbf{M}_{\mathbf{F}_0} \otimes \tilde{\mathbf{A}}_*) \mathcal{J}(\mathbf{I}_{R_0} \otimes \mathbf{S}) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix}, \end{aligned}$$

where $\mathbf{S} := (\mathbf{0}_{R_0 \times (T-R_0)}; \mathbf{I}_{(T-R_0)})$, $\mathcal{J}$ is an $R_0 T \times R_0 T$ commutation matrix, and

$$nT \mathbf{W}^{\frac{1}{2}} \mathbb{E} \left[\boldsymbol{\varphi}_{\text{RS}}(\boldsymbol{\beta}_0, \mathbf{F}_*) \boldsymbol{\varphi}_{\text{RS}}^\top(\boldsymbol{\beta}_0, \mathbf{F}_*) | \mathcal{D}_{nT} \right] \mathbf{W}^{\frac{1}{2}} = (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathcal{X}}^\top) \boldsymbol{\Sigma}_{\mathcal{D}} (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathcal{X}}) =: \boldsymbol{\Omega}.$$

Now, let

$$\begin{aligned} & nT (\mathbb{G}_{\text{RS}}^\top \mathbf{W} \mathbb{G}_{\text{RS}})^{-1} \mathbb{G}_{\text{RS}}^\top \mathbf{W} \mathbb{H}_{\text{RS}} \mathbf{W} \mathbb{G}_{\text{RS}} (\mathbb{G}_{\text{RS}}^\top \mathbf{W} \mathbb{G}_{\text{RS}})^{-1} \\ &= nT \begin{pmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}^\top \mathbf{B} \\ \mathbf{B}^\top \mathbf{A} & \mathbf{B}^\top \mathbf{B} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A}^\top \boldsymbol{\Omega} \mathbf{A} & \mathbf{A}^\top \boldsymbol{\Omega} \mathbf{B} \\ \mathbf{B}^\top \boldsymbol{\Omega} \mathbf{A} & \mathbf{B}^\top \boldsymbol{\Omega} \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}^\top \mathbf{B} \\ \mathbf{B}^\top \mathbf{A} & \mathbf{B}^\top \mathbf{B} \end{pmatrix}^{-1} \\ &=: nT \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^\top & \boldsymbol{\Psi}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{A}^\top \boldsymbol{\Omega} \mathbf{A} & \mathbf{A}^\top \boldsymbol{\Omega} \mathbf{B} \\ \mathbf{B}^\top \boldsymbol{\Omega} \mathbf{A} & \mathbf{B}^\top \boldsymbol{\Omega} \mathbf{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^\top & \boldsymbol{\Psi}_{22} \end{pmatrix}. \end{aligned} \tag{B.26}$$

Multiplying out, up to a factor of nT , the top-left block of the matrix in (B.26) is given by

$$\boldsymbol{\Psi}_{11} \mathbf{A}^\top \boldsymbol{\Omega} \mathbf{A} \boldsymbol{\Psi}_{11} + \boldsymbol{\Psi}_{12} \mathbf{B}^\top \boldsymbol{\Omega} \mathbf{A} \boldsymbol{\Psi}_{11} + \boldsymbol{\Psi}_{11} \mathbf{A}^\top \boldsymbol{\Omega} \mathbf{B} \boldsymbol{\Psi}_{12}^\top + \boldsymbol{\Psi}_{12} \mathbf{B}^\top \boldsymbol{\Omega} \mathbf{B} \boldsymbol{\Psi}_{12}^\top.$$

Using,

$$\begin{aligned} & \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^\top & \boldsymbol{\Psi}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}^\top \mathbf{B} \\ \mathbf{B}^\top \mathbf{A} & \mathbf{B}^\top \mathbf{B} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1} & -(\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{B} (\mathbf{B}^\top \mathbf{B})^{-1} \\ -(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{A} (\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1} & (\mathbf{B}^\top \mathbf{M}_A \mathbf{B})^{-1} \end{pmatrix}, \end{aligned} \tag{B.27}$$

which follows by

$$\det \left(\begin{pmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}^\top \mathbf{B} \\ \mathbf{B}^\top \mathbf{A} & \mathbf{B}^\top \mathbf{B} \end{pmatrix} \right) = \det(\mathbf{A}^\top \mathbf{A}) \det(\mathbf{B}^\top \mathbf{M}_A \mathbf{B}) = \det(\mathbf{B}^\top \mathbf{B}) \det(\mathbf{A}^\top \mathbf{M}_B \mathbf{A}),$$

using, for example, Fact 6.5.27 in [Bernstein \(2009\)](#), from which (B.27) then follows using standard expressions for the inverse of a partitioned matrix (see, e.g., Proposition 2.8.7 in [Bernstein \(2009\)](#)). Thus

$$\begin{aligned}\Psi_{11}\mathbf{A}^\top\Omega\mathbf{A}\Psi_{11} &= (\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1}\mathbf{A}^\top\Omega\mathbf{A}(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1} \\ \Psi_{12}\mathbf{B}^\top\Omega\mathbf{A}\Psi_{11} &= -(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{P}_B\Omega\mathbf{A}(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1} \\ \Psi_{11}\mathbf{A}^\top\Omega\mathbf{B}\Psi_{12}^\top &= -(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1}\mathbf{A}^\top\Omega\mathbf{P}_B\mathbf{A}(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1} \\ \Psi_{12}\mathbf{B}^\top\Omega\mathbf{B}\Psi_{12}^\top &= (\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{P}_B\Omega\mathbf{P}_B\mathbf{A}(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1},\end{aligned}$$

whereby

$$\begin{aligned}\Psi_{11}\mathbf{A}^\top\Omega\mathbf{A}\Psi_{11} + \Psi_{12}\mathbf{B}^\top\Omega\mathbf{A}\Psi_{11} + \Psi_{11}\mathbf{A}^\top\Omega\mathbf{B}\Psi_{12}^\top + \Psi_{12}\mathbf{B}^\top\Omega\mathbf{B}\Psi_{12}^\top \\ = (\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{M}_B\Omega\mathbf{M}_B\mathbf{A}(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1}.\end{aligned}$$

Now an expression for $\mathbf{M}_B$ is obtained. First,

$$\begin{aligned}\mathbf{B}^\top\mathbf{B} &= (\mathbf{I}_{R_0} \otimes \mathbf{S})^\top \mathcal{J}^\top(\mathbf{M}_{F_0} \otimes \tilde{\mathbf{\Lambda}}_*^\top \tilde{\mathbf{\Lambda}}_*) \mathcal{J}(\mathbf{I}_{R_0} \otimes \mathbf{S}) \\ &= (\mathbf{I}_{R_0} \otimes \mathbf{S})^\top (\tilde{\mathbf{\Lambda}}_*^\top \tilde{\mathbf{\Lambda}}_* \otimes \mathbf{M}_{F_0})(\mathbf{I}_{R_0} \otimes \mathbf{S}) \\ &= (\tilde{\mathbf{\Lambda}}_*^\top \tilde{\mathbf{\Lambda}}_* \otimes \mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S}).\end{aligned}$$

Notice that, since $\tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top = \tilde{\mathbf{\Lambda}}_* \mathbf{F}_*^\top$, there exists an $R_0 \times R_0$ invertible matrix $\mathbf{H}$ such that $\tilde{\mathbf{\Lambda}}_* = \tilde{\mathbf{\Lambda}}_0 \mathbf{H}^{-\top}$ from which it follows that $\tilde{\mathbf{\Lambda}}_*^\top \tilde{\mathbf{\Lambda}}_* = \mathbf{H}^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{H}^{-\top}$ is invertible w.p.a.1 under Assumption [AE\(ii\)](#). Moreover, using $\mathbf{F}_*^\top \mathbf{S} = \Phi_0^\top$ and $\mathbf{F}_*^\top \mathbf{F}_* = \mathbf{I}_{R_0} + \Phi_0^\top \Phi_0 \succ \mathbf{0}$,

$$\begin{aligned}\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S} &= \mathbf{S}^\top \mathbf{M}_{F_*} \mathbf{S} \\ &= \mathbf{I}_{T-R_0} - \Phi_0 (\mathbf{F}_*^\top \mathbf{F}_*)^{-1} \Phi_0^\top \\ &= \mathbf{I}_{T-R_0} - \Phi_0 (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} \Phi_0^\top \\ &= (\mathbf{I}_{T-R_0} + \Phi_0 \Phi_0^\top)^{-1} \succ \mathbf{0},\end{aligned}$$

where the last line follows by applying the Sherman-Morrison-Woodbury formula (see, e.g., Corollary 2.8.8 in [Bernstein \(2009\)](#)). Hence, $\mathbf{B}^\top \mathbf{B}$ is invertible w.p.a.1, and therefore

$$\begin{aligned}\mathbf{P}_B &= (\mathbf{M}_{F_0} \otimes \tilde{\mathbf{\Lambda}}_*) \mathcal{J}(\mathbf{I}_{R_0} \otimes \mathbf{S}) \left((\tilde{\mathbf{\Lambda}}_*^\top \tilde{\mathbf{\Lambda}}_*)^{-1} \otimes (\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \right) (\mathbf{I}_{R_0} \otimes \mathbf{S})^\top \mathcal{J}^\top(\mathbf{M}_{F_0} \otimes \tilde{\mathbf{\Lambda}}_*^\top) \\ &= (\mathbf{M}_{F_0} \otimes \tilde{\mathbf{\Lambda}}_*) \mathcal{J} \left((\tilde{\mathbf{\Lambda}}_*^\top \tilde{\mathbf{\Lambda}}_*)^{-1} \otimes \mathbf{S}(\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \mathbf{S}^\top \right) \mathcal{J}^\top(\mathbf{M}_{F_0} \otimes \tilde{\mathbf{\Lambda}}_*^\top) \\ &= (\mathbf{M}_{F_0} \otimes \tilde{\mathbf{\Lambda}}_*) \left(\mathbf{S}(\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \mathbf{S}^\top \otimes (\tilde{\mathbf{\Lambda}}_*^\top \tilde{\mathbf{\Lambda}}_*)^{-1} \right) (\mathbf{M}_{F_0} \otimes \tilde{\mathbf{\Lambda}}_*^\top) \\ &= (\mathbf{M}_{F_0} \mathbf{S}(\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{M}_{F_0} \otimes \mathbf{P}_{\tilde{\mathbf{\Lambda}}_0}).\end{aligned}\tag{B.28}$$

Consider $\mathbf{M}_{F_0} \mathbf{S} (\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{M}_{F_0}$. Under R-RS,

$$\begin{aligned} \mathbf{M}_{F_0} &= \mathbf{M}_{F_*} = \mathbf{I}_T - \begin{pmatrix} \mathbf{I}_{R_0} \\ \Phi_0 \end{pmatrix} \left(\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0 \right)^{-1} \begin{pmatrix} \mathbf{I}_{R_0} \\ \Phi_0 \end{pmatrix}^\top \\ &= \mathbf{I}_T - \begin{pmatrix} (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} & (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} \Phi_0^\top \\ \Phi_0 (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} & \Phi_0 (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} \Phi_0^\top \end{pmatrix}. \end{aligned}$$

Now, for brevity, let $\mathbf{G} := \mathbf{F}_*^\top \mathbf{F}_*$ whereby

$$\mathbf{M}_{F_*} \mathbf{S} = \begin{pmatrix} -\mathbf{G}^{-1} \Phi_0^\top \\ \mathbf{I}_{T-R_0} - \Phi_0 \mathbf{G}^{-1} \Phi_0^\top \end{pmatrix} = \begin{pmatrix} -\mathbf{G}^{-1} \Phi_0^\top \\ \mathbf{S}^\top \mathbf{M}_{F_*} \mathbf{S} \end{pmatrix},$$

and therefore

$$\mathbf{M}_{F_0} \mathbf{S} (\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{M}_{F_0} = \begin{pmatrix} \mathbf{G}^{-1} \Phi_0^\top (\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \Phi_0 \mathbf{G}^{-1} & -\mathbf{G}^{-1} \Phi_0^\top \\ -\Phi_0 \mathbf{G}^{-1} & \mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S} \end{pmatrix}.$$

Next, using $\Phi_0^\top \Phi_0 \mathbf{G}^{-1} = \mathbf{I}_{R_0} - \mathbf{G}^{-1}$ and $(\mathbf{S}^\top \mathbf{M}_{F_*} \mathbf{S})^{-1} = \mathbf{I}_{T-R_0} + \Phi_0 \Phi_0^\top$,

$$\begin{aligned} \mathbf{G}^{-1} \Phi_0^\top (\mathbf{S}^\top \mathbf{M}_{F_*} \mathbf{S})^{-1} \Phi_0 \mathbf{G}^{-1} &= \mathbf{G}^{-1} \Phi_0^\top \Phi_0 \mathbf{G}^{-1} + \mathbf{G}^{-1} \Phi_0^\top \Phi_0 \Phi_0^\top \Phi_0 \mathbf{G}^{-1} \\ &= \mathbf{G}^{-1} (\mathbf{I}_{R_0} - \mathbf{G}^{-1}) + (\mathbf{I}_{R_0} - \mathbf{G}^{-1}) (\mathbf{I}_{R_0} - \mathbf{G}^{-1}) \\ &= \mathbf{I}_{R_0} - \mathbf{G}^{-1}. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbf{M}_{F_0} \mathbf{S} (\mathbf{S}^\top \mathbf{M}_{F_0} \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{M}_{F_0} &= \begin{pmatrix} \mathbf{I}_{R_0} - \mathbf{G}^{-1} & -\mathbf{G}^{-1} \Phi_0^\top \\ -\Phi_0 \mathbf{G}^{-1} & \mathbf{I}_{T-R_0} - \Phi_0 \mathbf{G}^{-1} \Phi_0^\top \end{pmatrix} \\ &= \mathbf{I}_T - \begin{pmatrix} \mathbf{G}^{-1} & \mathbf{G}^{-1} \Phi_0^\top \\ \Phi_0 \mathbf{G}^{-1} & \Phi_0 \mathbf{G}^{-1} \Phi_0^\top \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} & (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} \Phi_0^\top \\ \Phi_0 (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} & \Phi_0 (\mathbf{I}_{R_0} + \Phi_0^\top \Phi_0)^{-1} \Phi_0^\top \end{pmatrix} \\ &= \mathbf{M}_{F_0}. \end{aligned}$$

Returning to (B.28), $\mathbf{P}_B = (\mathbf{M}_{F_0} \otimes \mathbf{P}_{\tilde{\Lambda}_0})$ from which it follows that $\mathbf{M}_B = \mathbf{I}_{T^2 K} - (\mathbf{M}_{F_0} \otimes \mathbf{P}_{\tilde{\Lambda}_0})$.

Applying this result

$$\begin{aligned} \mathbf{A}^\top \mathbf{M}_B \mathbf{A} &= \tilde{\mathbf{X}}^\top (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \tilde{\mathbf{X}} \\ \mathbf{A}^\top \mathbf{M}_B \Omega \mathbf{M}_B \mathbf{A} &= \tilde{\mathbf{X}}^\top (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \tilde{\Sigma}_{\mathcal{D}} (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \tilde{\mathbf{X}}. \end{aligned}$$

Thus,

$$\begin{aligned}
& nT(\Psi_{11}\mathbf{A}^\top\Omega\mathbf{A}\Psi_{11} + \Psi_{12}\mathbf{B}^\top\Omega\mathbf{A}\Psi_{11} + \Psi_{11}\mathbf{A}^\top\Omega\mathbf{B}\Psi_{12}^\top + \Psi_{12}\mathbf{B}^\top\Omega\mathbf{B}\Psi_{12}^\top) \\
&= (\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{M}_B\Omega\mathbf{M}_B\mathbf{A}(\mathbf{A}^\top\mathbf{M}_B\mathbf{A})^{-1} \\
&= \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1},
\end{aligned}$$

which concludes the result. $\square$

Proof of Proposition 6.

First write

$$\begin{aligned}
\sqrt{nT}\mathbf{W}^{\frac{1}{2}}\varphi_{\text{ALS}}(\beta, \mathbf{V}(\Phi)) &= \text{vec}(\mathbf{Q}_\mathcal{X}^\top(\mathbf{Y} - \mathbf{X} \cdot \beta)\mathbf{V}(\Phi)) \\
&= (\mathbf{I}_{T-R_0} \otimes \mathbf{Q}_\mathcal{X}^\top(\mathbf{Y} - \mathbf{X} \cdot \beta))\text{vec}(\mathbf{V}(\Phi)) \\
&= (\mathbf{I}_{T-R_0} \otimes \mathbf{Q}_\mathcal{X}^\top(\mathbf{Y} - \mathbf{X} \cdot \beta))\mathcal{I}_1\text{vec}(\Phi) \\
&\quad + (\mathbf{I}_{T-R_0} \otimes \mathbf{Q}_\mathcal{X}^\top(\mathbf{Y} - \mathbf{X} \cdot \beta))\mathcal{I}_2,
\end{aligned}$$

where $\mathcal{I}_1$ is a $T(T-R_0) \times R_0(T-R_0)$ matrix containing the columns of a $T(T-R_0)$ identity matrix which correspond to the Φ -elements of $\text{vec}(\mathbf{V}(\Phi))$, and $\mathcal{I}_2 := \text{vec}(\mathbf{V}(\Phi)) - \mathcal{I}_1\text{vec}(\Phi)$. Using this,

$$\begin{aligned}
\sqrt{nT}\mathbf{W}^{\frac{1}{2}}\mathbb{E}[\nabla\varphi_{\text{ALS}}(\beta_0, \mathbf{V}_*)|\mathcal{D}_{nT}] &= \begin{pmatrix} -(\mathbf{V}_* \otimes \mathbf{Q}_\mathcal{X})^\top \mathcal{X} & (\mathbf{I}_{T-R_0} \otimes \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top)\mathcal{I}_1 \end{pmatrix} \\
&=: \begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix},
\end{aligned}$$

and

$$nT\mathbf{W}^{\frac{1}{2}}\mathbb{E}\left[\varphi_{\text{ALS}}(\beta, \mathbf{V}_*)\varphi_{\text{ALS}}^\top(\beta, \mathbf{V}_*)\right]\mathbf{W}^{\frac{1}{2}} = (\mathbf{V}_* \otimes \mathbf{Q}_\mathcal{X})^\top \Sigma_{\mathcal{D}}(\mathbf{V}_* \otimes \mathbf{Q}_\mathcal{X}) =: \Omega.$$

Now,

$$\begin{aligned}
& nT(\mathbb{G}_{\text{ALS}}^\top \mathbf{W} \mathbb{G}_{\text{ALS}})^{-1} \mathbb{G}_{\text{ALS}}^\top \mathbf{W} \mathbb{H}_{\text{ALS}} \mathbf{W} \mathbb{G}_{\text{ALS}} (\mathbb{G}_{\text{ALS}}^\top \mathbf{W} \mathbb{G}_{\text{ALS}})^{-1} \\
&= nT \begin{pmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}^\top \mathbf{B} \\ \mathbf{B}^\top \mathbf{A} & \mathbf{B}^\top \mathbf{B} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A}^\top \Omega \mathbf{A} & \mathbf{A}^\top \Omega \mathbf{B} \\ \mathbf{B}^\top \Omega \mathbf{A} & \mathbf{B}^\top \Omega \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}^\top \mathbf{B} \\ \mathbf{B}^\top \mathbf{A} & \mathbf{B}^\top \mathbf{B} \end{pmatrix}^{-1} \\
&=: nT \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{pmatrix} \begin{pmatrix} \mathbf{A}^\top \Omega \mathbf{A} & \mathbf{A}^\top \Omega \mathbf{B} \\ \mathbf{B}^\top \Omega \mathbf{A} & \mathbf{B}^\top \Omega \mathbf{B} \end{pmatrix} \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{pmatrix}. \tag{B.29}
\end{aligned}$$

Multiplying out, up to a factor of nT , the top-left block of the matrix in (B.29) is given by

$$\Psi_{11}\mathbf{A}^\top\Omega\mathbf{A}\Psi_{11} + \Psi_{12}\mathbf{B}^\top\Omega\mathbf{A}\Psi_{11} + \Psi_{11}\mathbf{A}^\top\Omega\mathbf{B}\Psi_{12}^\top + \Psi_{12}\mathbf{B}^\top\Omega\mathbf{B}\Psi_{12}^\top.$$

Identically to (B.27),

$$\begin{aligned}
& \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}^\top \mathbf{B} \\ \mathbf{B}^\top \mathbf{A} & \mathbf{B}^\top \mathbf{B} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} (\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1} & -(\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{B} (\mathbf{B}^\top \mathbf{B})^{-1} \\ -(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{A} (\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1} & (\mathbf{B}^\top \mathbf{M}_A \mathbf{B})^{-1} \end{pmatrix}, \tag{B.30}
\end{aligned}$$

from which one obtains,

$$\begin{aligned}
& \Psi_{11} \mathbf{A}^\top \Omega \mathbf{A} \Psi_{11} + \Psi_{12} \mathbf{B}^\top \Omega \mathbf{A} \Psi_{11} + \Psi_{11} \mathbf{A}^\top \Omega \mathbf{B} \Psi_{12}^\top + \Psi_{12} \mathbf{B}^\top \Omega \mathbf{B} \Psi_{12}^\top \\
&= (\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{M}_B \Omega \mathbf{M}_B \mathbf{A} (\mathbf{A}^\top \mathbf{M}_B \mathbf{A})^{-1}.
\end{aligned}$$

An expression for $\mathbf{M}_B$ is now obtained. Notice $(\mathbf{I}_{T-R_0} \otimes \mathbf{F}_*^\top) \mathcal{I}_1 = \mathbf{I}_{(T-R_0)R_0}$. Thus,

$$\begin{aligned}
\mathbf{P}_B &= (\mathbf{I}_{T-R_0} \otimes \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top) \mathcal{I}_1 \left(\mathcal{I}_1^\top (\mathbf{I}_{T-R_0} \otimes \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top) \mathcal{I}_1 \right)^{-1} \mathcal{I}_1^\top (\mathbf{I}_{T-R_0} \otimes \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top) \\
&= (\mathbf{I}_{T-R_0} \otimes \tilde{\mathbf{A}}_* \mathbf{F}_*^\top) \mathcal{I}_1 \left(\mathcal{I}_1^\top (\mathbf{I}_{T-R_0} \otimes \mathbf{F}_* \tilde{\mathbf{A}}_*^\top \tilde{\mathbf{A}}_* \mathbf{F}_*^\top) \mathcal{I}_1 \right)^{-1} \mathcal{I}_1^\top (\mathbf{I}_{T-R_0} \otimes \mathbf{F}_* \tilde{\mathbf{A}}_*^\top) \\
&= (\mathbf{I}_{T-R_0} \otimes \mathbf{P}_{\tilde{\mathbf{A}}_0}),
\end{aligned}$$

and $\mathbf{M}_B = (\mathbf{I}_{T-R_0} \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0})$. Therefore, bringing together the above,

$$\begin{aligned}
& \Psi_{11} \mathbf{A}^\top \Omega \mathbf{A} \Psi_{11} + \Psi_{12} \mathbf{B}^\top \Omega \mathbf{A} \Psi_{11} + \Psi_{11} \mathbf{A}^\top \Omega \mathbf{B} \Psi_{12}^\top + \Psi_{12} \mathbf{B}^\top \Omega \mathbf{B} \Psi_{12}^\top \\
&= (\tilde{\mathcal{X}}^\top (\mathcal{V}_* \mathcal{V}_*^\top \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\mathcal{X}})^{-1} \tilde{\mathcal{X}}^\top (\mathcal{V}_* \mathcal{V}_*^\top \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\Sigma}_{\mathcal{D}} \\
&\quad \times (\mathcal{V}_* \mathcal{V}_*^\top \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\mathcal{X}} (\tilde{\mathcal{X}}^\top (\mathcal{V}_* \mathcal{V}_*^\top \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\mathcal{X}})^{-1},
\end{aligned}$$

from which the result then follows. $\square$

Proof of Proposition 7.

First, since $\mathcal{V}_*$ forms a basis for the left null space of $\mathbf{F}_0$, it follows that there exists a $(T-R_0) \times (T-R_0)$ matrix $\mathbf{H}$ such that $\mathcal{V}_* \mathbf{H} = \mathbf{U}$, where $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_{T-R_0}$ and $\mathbf{U} \mathbf{U}^\top = \mathbf{M}_{\mathbf{F}_0}$. Let

$$\begin{aligned}
\mathbf{A} &:= \frac{1}{\sqrt{nT}} (\mathcal{V}_*^\top \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\mathcal{X}} \\
\mathbf{B} &:= (\mathbf{H} \mathbf{H}^\top \otimes \mathbf{I}_{TK}).
\end{aligned}$$

Then

$$\begin{aligned}
\sigma_0^2 D^{-1} &= \sigma_0^2 \left(\frac{1}{nT} \tilde{\mathcal{X}}^\top (M_{F_0} \otimes M_{\tilde{\Lambda}_0}) \tilde{\mathcal{X}} \right)^{-1} \\
&= \sigma_0^2 \left(\frac{1}{nT} \tilde{\mathcal{X}}^\top (UU^\top \otimes M_{\tilde{\Lambda}_0}) \tilde{\mathcal{X}} \right)^{-1} \\
&= \sigma_0^2 \left(\frac{1}{nT} \tilde{\mathcal{X}}^\top (\mathbf{V}_* H H^\top \mathbf{V}_*^\top \otimes M_{\tilde{\Lambda}_0}) \tilde{\mathcal{X}} \right)^{-1} \\
&= \sigma_0^2 \left(\mathbf{A}^\top \mathbf{B} \mathbf{A} \right)^{-1}.
\end{aligned}$$

Let

$$\begin{aligned}
D_* &:= \frac{1}{nT} \tilde{\mathcal{X}}^\top (\mathbf{V}_* \mathbf{V}_*^\top \otimes M_{\tilde{\Lambda}_0}) \tilde{\mathcal{X}} \\
V_* &:= \frac{\sigma_0^2}{nT} \tilde{\mathcal{X}}^\top (\mathbf{V}_* \mathbf{V}_*^\top \mathbf{V}_* \mathbf{V}_*^\top \otimes M_{\tilde{\Lambda}_0}) \tilde{\mathcal{X}}.
\end{aligned}$$

Then

$$\begin{aligned}
D_*^{-1} V_* D_*^{-1} &= \sigma_0^2 \left(\mathbf{A}^\top \mathbf{A} \right)^{-1} \left(\mathbf{A}^\top (\mathbf{V}_*^\top \mathbf{V}_* \otimes I_{TK}) \mathbf{A} \right) \left(\mathbf{A}^\top \mathbf{A} \right)^{-1} \\
&= \sigma_0^2 \left(\mathbf{A}^\top \mathbf{A} \right)^{-1} \mathbf{A}^\top \mathbf{B}^{-1} \mathbf{A} \left(\mathbf{A}^\top \mathbf{A} \right)^{-1},
\end{aligned}$$

which follows because $\mathbf{V}_*^\top \mathbf{V}_* = \mathbf{H}^{-\top} \mathbf{U}^\top \mathbf{U} \mathbf{H}^{-1} = (\mathbf{H} \mathbf{H}^\top)^{-1}$. Now, let

$$\mathbf{C} := \begin{pmatrix} \mathbf{B} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{A}^\top \mathbf{B}^{-1} \mathbf{A} \end{pmatrix} = \begin{pmatrix} \mathbf{B}^{\frac{1}{2}} \\ \mathbf{A}^\top \mathbf{B}^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \mathbf{B}^{\frac{1}{2}} \\ \mathbf{A}^\top \mathbf{B}^{-\frac{1}{2}} \end{pmatrix}^\top,$$

which is positive semidefinite. Since $\mathbf{B}$ is positive definite, it then follows that $\mathbf{A}(\mathbf{A}^\top \mathbf{B}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \preceq \mathbf{B}$.⁴¹ Thus,

$$\sigma_0^2 (\mathbf{A}^\top \mathbf{B} \mathbf{A})^{-1} \preceq \sigma_0^2 (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{B}^{-1} \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1}.$$

□

Proof of Proposition 8.

See Appendix D.

Proof of Proposition 9.

See Appendix D.

⁴¹See, e.g., Fact 8.11.18 in Bernstein (2009).

Proof of Proposition 10.

To prove the result, the following three equations are demonstrated to hold:

$$\frac{\mu_{R_0}^*}{\mu_{R_0+1}^*} \xrightarrow{p} \infty. \quad (\text{B.31})$$

$$\frac{\mu_q^*}{\mu_{q+1}^*} = \mathcal{O}_p(1) \text{ for } q = 1, \dots, R_0 - 1, \quad (\text{B.32})$$

with $R_0 \geq 2$.

$$\frac{\mu_{R_0+s}^*}{\mu_{R_0+s+1}^*} = \mathcal{O}_p(1) \text{ for } s = 1, \dots, (T - R_0 - 1), \quad (\text{B.33})$$

with $T \geq R_0 + 2$. Thus, (B.31), (B.32), and (B.33), are now demonstrated, in reverse order.

Equation (B.33)

First, decompose

$$\begin{aligned} & \frac{1}{nT} \left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\beta} \right)^\top \left(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\beta} \right) \\ &= \frac{1}{nT} \left(\tilde{\mathbf{X}} \cdot (\beta_0 - \hat{\beta}) + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top + \tilde{\varepsilon} \right)^\top \left(\tilde{\mathbf{X}} \cdot (\beta_0 - \hat{\beta}) + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top + \tilde{\varepsilon} \right) \\ &=: \frac{1}{nT} \left(\mathbf{\Gamma} + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top \right)^\top \left(\mathbf{\Gamma} + \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top \right) \\ &=: \frac{1}{nT} \mathbf{F}_*^\top \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_*^\top + \frac{1}{nT} \mathbf{\Gamma}^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \mathbf{\Gamma} \\ &=: \mathbf{L}_1 + \mathbf{L}_2, \end{aligned} \quad (\text{B.34})$$

where $\mathbf{F}_* := \mathbf{F}_0 + \mathbf{\Gamma}^\top \tilde{\mathbf{\Lambda}}_0 (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1}$. Now,

$$0 < \mu_{R_0+s}^* = \mu_{R_0+s}(\mathbf{L}_1 + \mathbf{L}_2) + \varrho_n^2. \quad (\text{B.35})$$

Moreover, because both $\mathbf{L}_1$ and $\mathbf{L}_2$ are positive semidefinite, by Weyl's inequality (see, for example, Bernstein (2009) Theorem 8.4.11),

$$\mu_{R_0+s}(\mathbf{L}_2) \leq \mu_{R_0+s}(\mathbf{L}_1 + \mathbf{L}_2), \quad (\text{B.36})$$

and

$$\mu_{R_0+s}(\mathbf{L}_1 + \mathbf{L}_2) \leq \mu_{R_0+1}(\mathbf{L}_1) + \mu_s(\mathbf{L}_2) = \mu_s(\mathbf{L}_2), \quad (\text{B.37})$$

since $\mathbf{F}_*$ has R_0 columns. Hence, (B.35), (B.36), and (B.37) together imply

$$0 < \mu_{R_0+s}(\mathbf{L}_2) + \varrho_n^2 \leq \mu_{R_0+s}^* \leq \mu_s(\mathbf{L}_2) + \varrho_n^2, \quad (\text{B.38})$$

whereby

$$\frac{\mu_{R_0+s}^*}{\mu_{R_0+s+1}^*} \leq \frac{\mu_s \left(\frac{1}{nT} \mathbf{\Gamma}^\top \mathbf{M}_{\tilde{\mathbf{A}}_0} \mathbf{\Gamma} \right) + \varrho_n^2}{\mu_{R_0+s+1} \left(\frac{1}{nT} \mathbf{\Gamma}^\top \mathbf{M}_{\tilde{\mathbf{A}}_0} \mathbf{\Gamma} \right) + \varrho_n^2}. \quad (\text{B.39})$$

It remains only to show that the right-hand side of the above is $\mathcal{O}_p(1)$. Since $\mathbf{\Gamma}^\top \mathbf{M}_{\tilde{\mathbf{A}}_0} \mathbf{\Gamma}$ is positive semidefinite, for $r = 1, \dots, T$,

$$\begin{aligned} \mu_r(\mathbf{L}_2) &\leq \frac{1}{nT} \left\| \mathbf{M}_{\tilde{\mathbf{A}}_0} (\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\varepsilon}}) \right\|_2^2 \\ &\leq \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2^2 \left(\frac{1}{nT} \sum_{k=1}^K \|\tilde{\mathbf{X}}_k\|_F^2 \right) + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \left(\frac{1}{nT} \sum_{k=1}^K \|\tilde{\mathbf{X}}_k\|_F^2 \right)^{\frac{1}{2}} \frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}} + \frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2^2}{nT} \\ &= \mathcal{O}_p(r_{nT}^2), \end{aligned} \quad (\text{B.40})$$

which follows since $\|\tilde{\mathbf{X}}_k\|_F = \mathcal{O}_p(\sqrt{nT})$, and because $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 = \mathcal{O}_p(r_{nT})$ by assumption. Hence, $\mu_r(\mathbf{L}_2) = \mathcal{O}_p(r_{nT}^2)$, from which (B.33) then follows since $\varrho_n^{-1} r_{nT} = \mathcal{O}(1)$.

Equation (B.32)

Let $J \leq R_0$, $\mathbf{V}_J^*$ denote a $T \times J$ matrix of normalised eigenvectors associated with the J largest eigenvalues of $(nT)^{-1} \mathbf{F}_* \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_*^\top$, and, similarly, let $\mathbf{V}_J$ denote a $T \times J$ matrix of normalised eigenvectors associated with the J largest eigenvalues of $(nT)^{-1} \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top$.

$$\begin{aligned} \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_* \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_*^\top \right) &= \frac{1}{nT} \text{tr} \left(\mathbf{V}_J^{*\top} \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top \mathbf{V}_J^* \right) + \frac{2}{nT} \text{tr} \left(\mathbf{V}_J^{*\top} \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \mathbf{\Gamma} \mathbf{V}_J^* \right) \\ &\quad + \frac{1}{nT} \text{tr} \left(\mathbf{V}_J^{*\top} \mathbf{\Gamma}^\top \mathbf{P}_{\tilde{\mathbf{A}}_0} \mathbf{\Gamma} \mathbf{V}_J^* \right) \\ &\leq \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top \right) + \mathcal{O}_p(r_{nT}), \end{aligned} \quad (\text{B.41})$$

using $|\text{tr}(\mathbf{A})| \leq \text{rank}(\mathbf{A}) \|\mathbf{A}\|_2$, $\|\mathbf{V}_J^*\|_2 = \|\mathbf{V}_J\|_2 = 1$, $(nT)^{-\frac{1}{2}} \|\mathbf{\Gamma}\|_F = \mathcal{O}_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2) + (nT)^{-\frac{1}{2}} \mathcal{O}_p(\|\tilde{\boldsymbol{\varepsilon}}\|_2)$, and Lemma B.1(ii). Next,

$$\begin{aligned} \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_* \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_*^\top \right) &\geq \frac{1}{nT} \text{tr} \left(\mathbf{V}_J^\top \mathbf{F}_* \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_*^\top \mathbf{V}_J \right) \\ &= \frac{1}{nT} \text{tr} \left(\mathbf{V}_J^\top \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top \mathbf{V}_J \right) + \frac{2}{nT} \text{tr} \left(\mathbf{V}_J^\top \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \mathbf{\Gamma} \mathbf{V}_J \right) \\ &\quad + \frac{1}{nT} \text{tr} \left(\mathbf{V}_J^\top \mathbf{\Gamma}^\top \mathbf{P}_{\tilde{\mathbf{A}}_0} \mathbf{\Gamma} \mathbf{V}_J \right) \\ &= \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top \right) + \mathcal{O}_p(r_{nT}), \end{aligned} \quad (\text{B.42})$$

by similar steps to (B.41). This implies

$$\mu_j \left(\frac{1}{nT} \mathbf{F}_* \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_*^\top \right) = \mu_j \left(\frac{1}{nT} \mathbf{F}_0 \tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0 \mathbf{F}_0^\top \right) + \mathcal{O}_p(r_{nT}) \quad (\text{B.43})$$

since (B.41) and (B.42) hold for any $J \leq R_0$. Next, by Weyl's inequality,

$$\mu_q(\mathbf{L}_1 + \mathbf{L}_2) \leq \mu_q(\mathbf{L}_1) + \mu_1(\mathbf{L}_2),$$

whereby for any $J \leq R_0$,

$$\begin{aligned} \sum_{j=1}^J \mu_j(\mathbf{L}_1 + \mathbf{L}_2) &\leq \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_* \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_*^\top \right) + J \mu_1 \left(\frac{1}{nT} \mathbf{\Gamma}^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \mathbf{\Gamma} \right) \\ &= \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_0 \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top \right) + \mathcal{O}_p(r_{nT}), \end{aligned} \quad (\text{B.44})$$

using (B.40) and (B.43). Also,

$$\begin{aligned} \sum_{j=1}^J \mu_j(\mathbf{L}_1 + \mathbf{L}_2) &\geq \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_* \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_*^\top \right) + J \mu_T \left(\frac{1}{nT} \mathbf{\Gamma}^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \mathbf{\Gamma} \right) \\ &= \sum_{j=1}^J \mu_j \left(\frac{1}{nT} \mathbf{F}_0 \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top \right) + \mathcal{O}_p(r_{nT}), \end{aligned} \quad (\text{B.45})$$

again, using (B.40) and (B.43). Combining (B.44) and (B.45), and using the fact that $\mu_q(\mathbf{L}_1 + \mathbf{L}_2 + \varrho_n^2 \mathbf{I}_T) = \mu_q(\mathbf{L}_1 + \mathbf{L}_2) + \varrho_n^2$, gives the result

$$\mu_q(\mathbf{L}_1 + \mathbf{L}_2 + \varrho_n^2 \mathbf{I}_T) = \mu_q \left(\frac{1}{nT} \mathbf{F}_0 \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top \right) + \mathcal{O}_p(r_{nT}) + \varrho_n^2 \quad (\text{B.46})$$

for $q = 1, \dots, R_0$, where $0 < (nT)^{-1} \mu_q(\mathbf{F}_0 \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top)$ and $(nT)^{-1} \mu_q(\mathbf{F}_0 \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top) = \mathcal{O}_p(1)$ because first R_0 eigenvalues of the matrix $(nT)^{-1} \mathbf{F}_0 \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \mathbf{F}_0^\top$ equal those of $(n^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)(T^{-1} \mathbf{F}_0^\top \mathbf{F}_0)$ which, under Assumptions AE(ii) and AE(iii), converges in probability to a positive definite matrix with bounded eigenvalues. Equation (B.46) then establishes equation (B.32).

Equation (B.31)

Combining (B.38), (B.40), and (B.46) proves (B.31). □

Proof of Theorem 2.

See Appendix E.

Proof of Proposition 11.

First,

$$\begin{aligned}
\bar{D}_+ - \bar{D}_{+,LS} &= \frac{1}{nT} \tilde{\mathbf{H}}^\top (M_{F_0} \otimes M_{\tilde{\Lambda}_0}) \tilde{\mathbf{H}} - \frac{1}{nT} \tilde{\mathbf{H}}^\top (M_{F_0} \otimes M_{\Lambda_0}) \tilde{\mathbf{H}} \\
&\quad + \frac{\sigma_0^2}{nT} \text{tr}((\mathbf{G}\mathbf{G}^\top \otimes P_{\mathbf{X}})) \boldsymbol{\pi}_{K+1} \boldsymbol{\pi}_{K+1}^\top - \frac{\sigma_0^2}{nT} \text{tr}((\mathbf{G}\mathbf{G}^\top \otimes I_n)) \boldsymbol{\pi}_{K+1} \boldsymbol{\pi}_{K+1}^\top \\
&= \frac{1}{nT} \tilde{\mathbf{H}}^\top (M_{F_0} \otimes P_{\Lambda_0} - P_{P_{\mathbf{X}} \Lambda_0}) \tilde{\mathbf{H}} \\
&\quad + \frac{\sigma_0^2}{nT} \text{tr}((\mathbf{G}\mathbf{G}^\top \otimes -M_{\mathbf{X}})) \boldsymbol{\pi}_{K+1} \boldsymbol{\pi}_{K+1}^\top.
\end{aligned}$$

Following similar steps to those in the proof of Proposition 3, the first term in the above can be shown to be negative semidefinite. For the second term,

$$\frac{\sigma_0^2}{nT} \text{tr}(\mathbf{G}\mathbf{G}^\top \otimes -M_{\mathbf{X}}) = -\frac{\sigma_0^2}{T} \max\{0, 1 - \gamma K\} \text{tr}(\mathbf{G}\mathbf{G}^\top) \leq 0,$$

from which the result then follows.

Appendix C. Proofs of Lemmas B.1 - B.3

Intermediary Lemmas

Lemma C.1. Under Assumptions MD, ED, CS, and AE,

- (i) $\|\mathbf{F}_0^\top \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{X}}_k\|_2 = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K$;
- (ii) $\|\tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p(\sqrt{nT})$;
- (iii) $\|\tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_2 = \mathcal{O}_p(\sqrt{nT})$;
- (iv) $\|\mathbf{P}_{F_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\Lambda}_0}\|_2 = \mathcal{O}_p(1)$;
- (v) $\|\tilde{\mathbf{X}}_k \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\Lambda}_0\|_2 = \mathcal{O}_p(n\sqrt{T})$ for $k = 1, \dots, K$;
- (vi) $\|\tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_2 = \mathcal{O}_p(T)$;
- (vii) $\|\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{X}}_k\|_2 = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K$;
- (viii) $\|\tilde{\boldsymbol{\varepsilon}} \tilde{\mathbf{X}}_k^\top\|_2 = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K$.

Lemma C.2. Under Assumptions MD, ED, CS, AE, and AD,

- (i) $(nT)^{-\frac{1}{2}} \text{tr}(\tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{B} \tilde{\boldsymbol{\varepsilon}}) = \psi_k^{(1)} + \mathcal{O}_p(n^{-\frac{1}{4}})$;
- (ii) $(nT)^{-\frac{1}{2}} \text{tr}(\mathbf{M}_{F_0} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{B} \tilde{\mathbf{X}}_k) = \psi_k^{(2)} + \mathcal{O}_p(n^{-\frac{1}{2}})$,

for $k = 1, \dots, K$, where $\mathbf{B} := \mathbf{F}_0(\mathbf{F}_0^\top \mathbf{F}_0)^{-1}(\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top$.

Proof of Lemma B.1(i). Given $\|\tilde{\mathbf{X}}_k\|_2 = \|\mathbf{Q}_{\mathbf{X}}^\top \mathbf{X}_k\|_2 \leq \|\mathbf{X}_k\|_2 \leq \|\mathbf{X}_k\|_F$ since $\|\mathbf{Q}_{\mathbf{X}}\|_2 = 1$, it suffices to show that $\|\mathbf{X}_k\|_F = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K$. Now,

$$\mathbb{E}[\|\mathbf{X}_k\|_F^2] = \mathbb{E}\left[\sum_{i=1}^n \sum_{t=1}^T x_{kit}^2\right] = \mathcal{O}(nT),$$

by Assumption **MD(ii)**, which gives the result $\|\mathbf{X}_k\|_F = \mathcal{O}_p(\sqrt{nT})$ using Markov's inequality. $\square$

Proof of Lemma B.1(ii). For the first part,

$$\mathbb{E} [\|\mathbf{F}_0\|_F^2] = \mathbb{E} \left[\sum_{t=1}^T \sum_{r=1}^{R_0} f_{0,tr}^2 \right] = \mathcal{O}(T), \quad (\text{C.1})$$

using Assumption **MD(ii)**. This then implies $\|\mathbf{F}_0\|_F = \mathcal{O}_p(\sqrt{T})$. For the second part, since $\|\tilde{\mathbf{A}}_0\|_F = \|\mathbf{Q}_X^\top \mathbf{A}_0\|_F \leq \|\mathbf{A}_0\|_F$, it suffices to show that $\mathbb{E}[\|\mathbf{A}_0\|_F^2] = \mathcal{O}(n)$, which follows by the same steps as (C.1) using Assumption **MD(ii)**. The results for the spectral norm follow from the inequality $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$. $\square$

Proof of Lemma B.2(i). The lemma is obtained by combining two results. The first is that

$$\|\tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p(T^{\frac{3}{4}}), \quad (\text{C.2})$$

and the second is that

$$\|\tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p \left(\sqrt{n} T^{\frac{1}{4}} + n^{\frac{1}{4}} \sqrt{T} \right). \quad (\text{C.3})$$

Taken together one obtains the result.

Expression (C.2)

To establish (C.2), first,

$$\|\tilde{\boldsymbol{\varepsilon}}\|_2^2 = \|\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}}\|_2 \leq \|\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}}\|_F = \text{tr}(\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}})^{\frac{1}{2}}.$$

Now,

$$\begin{aligned} \mathbb{E} \left[\text{tr}(\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}}) \right] &= \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n \mathbb{E} \left[(\mathbf{P}\mathbf{x})_{i_1 i_2} (\mathbf{P}\mathbf{x})_{i_3 i_4} \boldsymbol{\varepsilon}_{i_1}^\top \boldsymbol{\varepsilon}_{i_4} \boldsymbol{\varepsilon}_{i_2}^\top \boldsymbol{\varepsilon}_{i_3} \right] \\ &= \sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \mathbb{E} [(\mathbf{P}\mathbf{x})_{i_1 i_1} (\mathbf{P}\mathbf{x})_{i_2 i_2} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{C}, i_1} \boldsymbol{\Sigma}_{\mathcal{C}, i_2})] \\ &\quad + \sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \mathbb{E} [(\mathbf{P}\mathbf{x})_{i_1 i_2}^2 \text{tr}(\boldsymbol{\Sigma}_{\mathcal{C}, i_1} \boldsymbol{\Sigma}_{\mathcal{C}, i_2})] \\ &\quad + \sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \mathbb{E} [(\mathbf{P}\mathbf{x})_{i_1 i_2} (\mathbf{P}\mathbf{x})_{i_2 i_1} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{C}, i_1}) \text{tr}(\boldsymbol{\Sigma}_{\mathcal{C}, i_2})] \\ &\quad + \sum_{i=1}^n \mathbb{E} \left[(\mathbf{P}\mathbf{x})_{ii}^2 \boldsymbol{\varepsilon}_i^\top \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \boldsymbol{\varepsilon}_i \right] \\ &=: p_1 + p_2 + p_3 + p_4. \end{aligned}$$

Consider each of these terms.

$$\begin{aligned}
|p_1| &\leq \sum_{i_1=1}^n \sum_{i_2=1}^n \mathbb{E} [(\mathbf{P}\mathbf{x})_{i_1 i_1} (\mathbf{P}\mathbf{x})_{i_2 i_2} |\text{tr}(\mathbf{\Sigma}_{\mathcal{C}, i_1} \mathbf{\Sigma}_{\mathcal{C}, i_2})|] \lesssim T (\mathbb{E}[\text{tr}(\mathbf{P}\mathbf{x})^2]) = \mathcal{O}(T^3) \\
|p_2| &\leq \sum_{i_1=1}^n \sum_{i_2=1}^n \mathbb{E} [(\mathbf{P}\mathbf{x})_{i_1 i_2}^2 |\text{tr}(\mathbf{\Sigma}_{\mathcal{C}, i_1} \mathbf{\Sigma}_{\mathcal{C}, i_2})|] \lesssim T \mathbb{E} [\|\mathbf{P}\mathbf{x}\|_F^2] = \mathcal{O}(T^2) \\
|p_3| &\leq \sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \mathbb{E} [(\mathbf{P}\mathbf{x})_{i_1 i_2} (\mathbf{P}\mathbf{x})_{i_2 i_1} |\text{tr}(\mathbf{\Sigma}_{\mathcal{C}, i_1})| |\text{tr}(\mathbf{\Sigma}_{\mathcal{C}, i_2})|] \lesssim T^2 \mathbb{E} [\|\mathbf{P}\mathbf{x}\|_F^2] = \mathcal{O}(T^3),
\end{aligned}$$

and

$$|p_4| \leq \sum_{i=1}^n \mathbb{E} \left[(\mathbf{P}\mathbf{x})_{ii}^2 \sum_{t_1=1}^T \sum_{t_2=1}^T \varepsilon_{it_1}^2 \varepsilon_{it_2}^2 \right] \lesssim T^2 \sum_{i=1}^n \mathbb{E} [(\mathbf{P}\mathbf{x})_{ii}^2] \leq T^2 \mathbb{E} [\|\mathbf{P}\mathbf{x}\|_F^2] = \mathcal{O}(T^3),$$

which follows using Assumption [EC](#). Thus, $\mathbb{E} [\text{tr}(\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}})] = \mathcal{O}(T^3)$, from which it then follows that $\text{tr}(\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}})^{\frac{1}{2}} = \mathcal{O}_p(T^{\frac{3}{2}})$, and thereby $\|\tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p(T^{\frac{3}{4}})$. Notice that same conclusion continues to hold if Assumption [EC](#) is replaced by Assumption [ED](#).

Expression [\(C.3\)](#)

Let $\mathbf{u}_i = \boldsymbol{\varepsilon}_i / \sqrt{T}$. Under Assumption [EC](#),

$$\mathbb{E} \left[\left\| \mathbf{u}_i \mathbf{u}_i^\top \right\|_F^2 \right] = \mathbb{E} \left[\text{tr}(\mathbf{u}_i \mathbf{u}_i^\top \mathbf{u}_i \mathbf{u}_i^\top) \right] \leq \sum_{t_1=1}^T \sum_{t_2=1}^T (\mathbb{E}[u_{it_1}^4])^{\frac{1}{2}} (\mathbb{E}[u_{it_2}^4])^{\frac{1}{2}} \leq c.$$

Thus the Rosenthal inequality (Corollary 7.4) in [Mackey et al. \(2014\)](#) is applicable with $p = 1$ from which one obtains

$$\mathbb{E} \left[\left\| \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i^\top | \mathcal{C}_{nT} \right\|_F^2 \right] \leq \left(\left\| \sum_{i=1}^n \mathbb{E} [\mathbf{u}_i \mathbf{u}_i^\top | \mathcal{C}_{nT}] \right\|_F^{\frac{1}{2}} + \sqrt{2} \left(\sum_{i=1}^n \mathbb{E} \left[\left\| \mathbf{u}_i \mathbf{u}_i^\top \right\|_F^2 | \mathcal{C}_{nT} \right] \right)^{\frac{1}{4}} \right)^4.$$

Now,

$$\left\| \sum_{i=1}^n \mathbb{E} [\mathbf{u}_i \mathbf{u}_i^\top | \mathcal{C}_{nT}] \right\|_F^2 = \frac{1}{T^2} \left\| \sum_{i=1}^n \mathbf{\Sigma}_{\mathcal{C}, i} \right\|_F^2 = \frac{1}{T^2} \sum_{i_1=1}^n \sum_{i_2=1}^n \text{tr}(\mathbf{\Sigma}_{\mathcal{C}, i_1} \mathbf{\Sigma}_{\mathcal{C}, i_2}) \lesssim \frac{n^2}{T},$$

and

$$\sum_{i=1}^n \mathbb{E} \left[\left\| \mathbf{u}_i \mathbf{u}_i^\top \right\|_F^2 | \mathcal{C}_{nT} \right] = \frac{1}{T^2} \sum_{i=1}^n \mathbb{E} \left[\sum_{t_1=1}^T \sum_{t_2=1}^T \varepsilon_{it_1}^2 \varepsilon_{it_2}^2 | \mathcal{C}_{nT} \right] \lesssim n,$$

under Assumption [EC](#). Thus,

$$\mathbb{E} [\|\mathbf{u}\|_2^4] = \mathbb{E} [\|\mathbf{u}^\top \mathbf{u}\|_2^2] = \mathbb{E} \left[\left\| \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i^\top \right\|_2^2 \right] \leq \mathbb{E} \left[\left\| \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i^\top \right\|_F^2 \right] \lesssim \left(\frac{\sqrt{n}}{T^{\frac{1}{4}}} + \sqrt{2} n^{\frac{1}{4}} \right)^4,$$

from which it follows that

$$\|\boldsymbol{\varepsilon}\|_2 = \sqrt{T} \|\mathbf{u}\|_2 = \mathcal{O}_p \left(\sqrt{n} T^{\frac{1}{4}} + n^{\frac{1}{4}} \sqrt{T} \right),$$

which establishes (C.3). Notice that same conclusion continues to hold if Assumption EC is replaced by Assumption ED. $\square$

Proof of Lemma B.2(ii). The result follows from

$$\begin{aligned}
\mathbb{E} \left[\text{tr}(\tilde{\mathbf{X}}_k^\top \tilde{\boldsymbol{\varepsilon}})^2 \right] &= \mathbb{E} \left[\text{tr}(\mathbf{X}_k^\top \mathbf{P}_{\mathbf{X}} \boldsymbol{\varepsilon})^2 \right] \\
&= \mathbb{E} \left[\text{tr}(\mathbf{X}_k^\top \boldsymbol{\varepsilon})^2 \right] \\
&= \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T x_{ki_1 t_1} x_{ki_2 t_2} \varepsilon_{i_1 t_1} \varepsilon_{i_2 t_2} \right] \\
&\lesssim \mathbb{E} \left[\sum_{i_1=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T |x_{ki_1 t_1}| |x_{ki_1 t_2}| \right] \\
&= \mathcal{O}(nT^2),
\end{aligned}$$

using Assumption MD(ii), and either Assumption EC or Assumption ED. The result then follows using Markov's inequality. $\square$

Proof of Lemma B.3(i). Recall

$$\mathbf{b}^{(1)} = -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_1 \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\ \vdots \\ \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_K \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \end{pmatrix}. \quad (\text{C.4})$$

Now,

$$\begin{aligned}
b_k^{(1)} &= -\frac{1}{\sqrt{nT}} \text{tr} \left(\tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\
&\quad + \frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{X}}_k \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\
&\quad - \frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\
&=: b_k^{(1.1)} + b_k^{(1.2)} + b_k^{(1.3)}.
\end{aligned}$$

Consider each of these terms.

$$\begin{aligned}
|b_k^{(1.2)}| &\lesssim \frac{1}{\sqrt{nT}} \frac{1}{nT} \frac{1}{T} \|\mathbf{F}_0\|_2^2 \|\mathbf{F}_0^\top \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{X}}_k\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2^2 \left\| \left(\frac{1}{n} \tilde{\Lambda}_0^\top \tilde{\Lambda}_0 \right)^{-1} \right\|_2 \|\tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_2 \\
&= \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right),
\end{aligned}$$

and

$$\begin{aligned}
|b_k^{(1.3)}| &\lesssim \frac{1}{\sqrt{nT}} \frac{1}{nT} \|\mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\Lambda}_0}\|_2 \|\tilde{\mathbf{X}}_k\|_2 \|\mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\Lambda}_0^\top \tilde{\Lambda}_0 \right)^{-1} \right\|_2 \|\tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_2 \\
&= \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right),
\end{aligned}$$

using Lemmas B.1(i), B.1(ii), C.1(i), C.1(ii), and C.1(iv), as well as Assumption AE. Next, using Lemma C.2(i),

$$b_k^{(1.1)} = -\psi_k^{(1)} + \mathcal{O}_p\left(n^{-\frac{1}{4}}\right),$$

where

$$\begin{aligned} |\psi_k^{(1)}| &\lesssim \sqrt{\frac{T}{n}} \frac{1}{nT} \|\Sigma_{\mathcal{D}}\|_2 \|\tilde{\mathbf{X}}_k\|_2 \|\mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\tilde{\mathbf{\Lambda}}_0\|_2 \\ &= \mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right), \end{aligned}$$

using Lemmas B.1(i) and B.1(ii), as well as Assumption AE. □

Proof of Lemma B.3(ii). Recall

$$\mathbf{b}^{(2)} = -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}\left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{X}}_1^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\right) \\ \vdots \\ \text{tr}\left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{X}}_K^\top \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\right) \end{pmatrix}. \quad (\text{C.5})$$

First,

$$\begin{aligned} |b_k^{(2)}| &\leq \frac{1}{\sqrt{nT}} |\text{tr}(\tilde{\mathbf{X}}_k^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}})| \\ &\quad + \frac{1}{\sqrt{nT}} |\text{tr}(\mathbf{P}_{\mathbf{F}_0} \tilde{\mathbf{X}}_k^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}})| \\ &\quad + \frac{1}{\sqrt{nT}} |\text{tr}(\tilde{\mathbf{X}}_k^\top \mathbf{P}_{\tilde{\mathbf{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}})| \\ &\quad + \frac{1}{\sqrt{nT}} |\text{tr}(\mathbf{P}_{\mathbf{F}_0} \tilde{\mathbf{X}}_k^\top \mathbf{P}_{\tilde{\mathbf{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0)^{-1} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}})| \\ &=: b_k^{(2.1)} + b_k^{(2.2)} + b_k^{(2.3)} + b_k^{(2.4)}. \end{aligned}$$

Now,

$$b_k^{(2.1)} \lesssim \frac{1}{\sqrt{nT}} \frac{1}{nT} \|\tilde{\mathbf{X}}_k^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),$$

and

$$\begin{aligned} b_k^{(2.3)} &\lesssim \frac{1}{\sqrt{nT}} \frac{1}{n^2 T} \|\tilde{\mathbf{X}}_k\|_2 \|\tilde{\mathbf{\Lambda}}_0\|_2 \|\tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2^2 \|\tilde{\mathbf{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_2 \\ &= \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

using Lemmas B.1(i), B.1(ii), C.1(i), C.1(ii), C.1(iii), and C.1(iv), as well as Assumption AE. Similarly for $b_k^{(2.2)}$ and $b_k^{(2.4)}$ which gives the result. □

Proof of Lemma B.3(iii). Recall

$$\mathbf{b}^{(3)} = -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\mathbf{X}}_1) \\ \vdots \\ \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\mathbf{X}}_K) \end{pmatrix}. \quad (\text{C.6})$$

First,

$$\begin{aligned} b_k^{(3)} &= -\frac{1}{\sqrt{nT}} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\mathbf{X}}_k) \\ &\quad + \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\mathbf{X}}_k) \\ &\quad - \frac{1}{\sqrt{nT}} \text{tr}(\mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\mathbf{X}}_k) \\ &=: b_k^{(3.1)} + b_k^{(3.2)} + b_k^{(3.3)}. \end{aligned}$$

For the second and third terms,

$$\begin{aligned} |b_k^{(3.2)}| &\lesssim \frac{1}{\sqrt{nT}} \frac{1}{n^2 T} \|\tilde{\mathbf{X}}_k \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\Lambda}}_0\|_2 \|\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \right)^{-1} \right\|_2^2 \|\tilde{\boldsymbol{\Lambda}}_0\|_2 \\ &= \mathcal{O}_p \left(\frac{1}{\sqrt{nT}} \right), \end{aligned}$$

and

$$\begin{aligned} |b_k^{(3.3)}| &\lesssim \frac{1}{\sqrt{nT}} \frac{1}{nT} \|\mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\boldsymbol{\Lambda}}_0}\|_2 \|\tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \right)^{-1} \right\|_2 \|\tilde{\boldsymbol{\Lambda}}_0\|_2 \|\tilde{\mathbf{X}}_k\|_2 \\ &= \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right), \end{aligned}$$

using Lemmas B.1(i), B.1(ii), C.1(iii), C.1(iv), C.1(v), and C.1(vi), as well as Assumption AE. For term $b_k^{(3.1)}$, using Lemma C.2(ii),

$$b_k^{(3.1)} = -\psi_k^{(2)} + \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right),$$

where

$$\begin{aligned} |\psi_k^{(2)}| &\lesssim \sqrt{\frac{T}{n}} \frac{1}{nT} \|\boldsymbol{\Sigma}_{\mathcal{D}}\|_2 \|\tilde{\mathbf{X}}_k\|_2 \|\mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \right)^{-1} \right\|_2 \|\tilde{\boldsymbol{\Lambda}}_0\|_2 \\ &= \mathcal{O}_p \left(\sqrt{\frac{T}{n}} \right), \end{aligned}$$

using Lemmas B.1(i) and B.1(ii), as well as Assumptions AE and AD. $\square$

Proofs of Intermediary Lemmas C.1 - C.2

Proof of Lemma C.1(i).

$$\begin{aligned}
\mathbb{E} \left[\|\mathbf{F}_0^\top \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{X}}_k\|_F^2 \right] &= \mathbb{E} \left[\|\mathbf{F}_0^\top \boldsymbol{\varepsilon}^\top \mathbf{X}_k\|_F^2 \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n (\mathbf{X}_k \mathbf{X}_k^\top)_{ii} \text{tr}(\mathbf{F}_0 \mathbf{F}_0^\top \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \right] \\
&\lesssim \mathbb{E} \left[\|\mathbf{X}_k\|_F^2 \|\mathbf{F}_0\|_F^2 \right] \\
&= \mathcal{O}(nT^2),
\end{aligned}$$

which follows by Assumption ED, using the Cauchy-Schwarz inequality, and noting that $\mathbb{E}[\|\mathbf{X}_k\|_F^4] = \mathcal{O}(n^2T^2)$ and $\mathbb{E}[\|\mathbf{F}_0\|_F^4] = \mathcal{O}(T^2)$ under Assumption MD(ii). $\square$

Proof of Lemma C.1(ii).

$$\begin{aligned}
\mathbb{E} \left[\|\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_F^2 \right] &= \mathbb{E} \left[\|(\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0)^\top \boldsymbol{\varepsilon}\|_F^2 \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0 \boldsymbol{\Lambda}_0^\top \mathbf{P}_{\mathcal{X}})_{ii} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, ii}) \right] \\
&\lesssim T \mathbb{E} \left[\|\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0\|_F^2 \right] \\
&\lesssim T \mathbb{E} \left[\|\boldsymbol{\Lambda}_0\|_F^2 \right] = \mathcal{O}(nT),
\end{aligned}$$

under Assumptions MD(ii) and ED. $\square$

Proof of Lemma C.1(iii).

$$\begin{aligned}
\mathbb{E} \left[\|\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_F^2 \right] &= \mathbb{E} \left[\|(\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0)^\top \boldsymbol{\varepsilon} \mathbf{F}_0\|_F^2 \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0 \boldsymbol{\Lambda}_0^\top \mathbf{P}_{\mathcal{X}})_{ii} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, ii} \mathbf{F}_0 \mathbf{F}_0^\top) \right] \\
&\lesssim \mathbb{E} \left[\|\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0\|_F^2 \|\mathbf{F}_0\|_F^2 \right] \\
&\lesssim \mathbb{E} \left[\|\boldsymbol{\Lambda}_0\|_F^2 \|\mathbf{F}_0\|_F^2 \right] = \mathcal{O}(nT),
\end{aligned}$$

under Assumptions MD(ii) and ED. $\square$

Proof of Lemma C.1(iv).

$$\begin{aligned}
\|\mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\boldsymbol{\Lambda}}_0}\|_2 &\leq \left\| \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\Lambda}}_0 (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \right\|_2 \\
&\leq \frac{1}{nT} \|\mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \mathbf{F}_0^\top \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\Lambda}}_0 \right\|_2 \left\| \left(\frac{1}{n} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \right)^{-1} \right\|_2 \|\tilde{\boldsymbol{\Lambda}}_0\|_2 = \mathcal{O}_p(1),
\end{aligned}$$

using Lemmas B.1(ii) and C.1(iii), as well as Assumption AE(ii) and AE(iii). $\square$

Proof of Lemma C.1(v).

$$\begin{aligned}
\mathbb{E} \left[\|\tilde{\mathbf{X}}_k \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{\Lambda}}_0\|_F^2 \right] &= \mathbb{E} \left[\|\mathbf{X}_k \boldsymbol{\varepsilon}^\top \mathbf{P}_{\mathcal{X}} \mathbf{\Lambda}_0\|_F^2 \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}} \mathbf{\Lambda}_0 \mathbf{\Lambda}_0^\top \mathbf{P}_{\mathcal{X}})_{ii} \text{tr}(\mathbf{X}_k^\top \mathbf{X}_k \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \right] \\
&\lesssim \mathbb{E} \left[\|\mathbf{P}_{\mathcal{X}} \mathbf{\Lambda}_0\|_F^2 \|\mathbf{X}_k\|_F^2 \right] \\
&\leq \mathbb{E} \left[\|\mathbf{\Lambda}_0\|_F^2 \|\mathbf{X}_k\|_F^2 \right] = \mathcal{O}(n^2 T),
\end{aligned}$$

which follows by Assumption **ED**, using the Cauchy-Schwarz inequality, and noting that $\mathbb{E}[\|\mathbf{X}_k\|_F^4] = \mathcal{O}(n^2 T^2)$ and $\mathbb{E}[\|\mathbf{\Lambda}_0\|_F^4] = \mathcal{O}(n^2)$ under Assumption **MD(ii)**. $\square$

Proof of Lemma C.1(vi).

$$\begin{aligned}
\mathbb{E} \left[\|\tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_F^2 \right] &= \mathbb{E} \left[\|\mathbf{P}_{\mathcal{X}} \boldsymbol{\varepsilon} \mathbf{F}_0\|_F^2 \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}})_{ii} \text{tr}(\mathbf{F}_0 \mathbf{F}_0^\top \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \right] \\
&\lesssim T \mathbb{E} \left[\|\mathbf{F}_0\|_F^2 \right] = \mathcal{O}(T^2),
\end{aligned}$$

which follows by Assumptions **MD(ii)** and **ED**. $\square$

Proof of Lemma C.1(vii).

$$\begin{aligned}
\mathbb{E} \left[\|\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{X}}_k\|_F^2 \right] &= \mathbb{E} \left[\|\boldsymbol{\varepsilon}^\top \mathbf{X}_k\|_F^2 \right] \\
&= \mathbb{E} \left[\sum_{t_1=1}^T \sum_{t_2=1}^T \sum_{i_1=1}^n \sum_{i_2=1}^n \varepsilon_{i_1 t_2} x_{k i_1 t_1} \varepsilon_{i_2 t_2} x_{k i_2 t_1} \right] \\
&= \mathbb{E} \left[\sum_{t_1=1}^T \sum_{t_2=1}^T \sum_{i=1}^n \varepsilon_{i t_2}^2 x_{k i t_1}^2 \right] = \mathcal{O}(n T^2),
\end{aligned}$$

which follows by Assumptions **MD(ii)** and **ED**. $\square$

Proof of Lemma C.1(viii).

$$\begin{aligned}
\mathbb{E} \left[\|\tilde{\boldsymbol{\varepsilon}} \tilde{\mathbf{X}}_k^\top\|_F^2 \right] &= \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2=1}^n (\mathbf{P}_{\mathcal{X}})_{i_1 i_2} \text{tr}(\mathbf{X}_k^\top \mathbf{X}_k \boldsymbol{\varepsilon}_{i_1} \boldsymbol{\varepsilon}_{i_2}^\top) \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}})_{ii} \text{tr}(\mathbf{X}_k^\top \mathbf{X}_k \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \right] \\
&\lesssim T \mathbb{E} \left[\|\mathbf{X}_k\|_F^2 \right] = \mathcal{O}(n T^2),
\end{aligned}$$

which follows by Assumptions **MD(ii)** and **ED**. $\square$

Proof of Lemma C.2(i). Write,

$$\begin{aligned}
\frac{1}{\sqrt{nT}} \text{tr}(\tilde{\varepsilon}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{B} \tilde{\varepsilon}) &= \frac{1}{\sqrt{nT}} \text{vec}(\varepsilon)^\top (\mathbf{I}_T \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{B} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\varepsilon) \\
&= \frac{1}{\sqrt{nT}} \text{tr} \left(\Sigma_{\mathcal{D}} (\mathbf{I}_T \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{B} \mathbf{Q}_{\mathcal{X}}^\top) \right) \\
&\quad + \frac{1}{\sqrt{nT}} \text{tr} \left(\left(\text{vec}(\varepsilon) \text{vec}(\varepsilon)^\top - \Sigma_{\mathcal{D}} \right) (\mathbf{I}_T \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{B} \mathbf{Q}_{\mathcal{X}}^\top) \right). \quad (\text{C.7})
\end{aligned}$$

It is now demonstrated that the second term in the above is $\mathcal{O}_p(n^{-\frac{1}{4}})$. Let

$$\mathbf{\Pi} := \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \left(\frac{1}{n} \tilde{\Lambda}_0^\top \tilde{\Lambda}_0 \right)^{-1}.$$

One can write the second term on the right in (C.7) as

$$\begin{aligned}
&\sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \mathbf{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i_1=1}^n \sum_{i_2=1}^n \text{tr}(\varepsilon_{i_1} \varepsilon_{i_2}^\top - \Sigma_{\mathcal{D}, i_1 i_2}) (\mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{X}}_k \mathbf{f}_{0, r_1} \tilde{\Lambda}_{0, r_2}^\top \mathbf{Q}_{\mathcal{X}}^\top)_{i_1 i_2} \right) \\
&=: \sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \mathbf{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i_1=1}^n \sum_{i_2=1}^n \text{tr}(\varepsilon_{i_1} \varepsilon_{i_2}^\top - \Sigma_{\mathcal{D}, i_1 i_2}) \mathbf{A}_{i_1 i_2}(r_1, r_2) \right) \\
&= \sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \mathbf{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i=1}^n \text{tr}(\varepsilon_i \varepsilon_i^\top - \Sigma_{\mathcal{D}, ii}) \mathbf{A}_{ii}(r_1, r_2) \right) \\
&\quad + \sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \mathbf{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \text{tr}(\varepsilon_{i_1} \varepsilon_{i_2}^\top) \mathbf{A}_{i_1 i_2}(r_1, r_2) \right) \\
&=: p_1 + p_2.
\end{aligned}$$

Consider each of these terms, in turn. First, since $\|\mathbf{\Pi}\|_{\max} \leq \|\mathbf{\Pi}\|_2 = \mathcal{O}_p(1)$ under Assumption **AE(ii)** and **AE(iii)**, it suffices to examine the summands (over R_0) in p_1 and p_2 for each (r_1, r_2) . Starting with

p_1 , for brevity let $\mathbf{X}_k^* := \mathbf{Q}_\mathbf{X} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_\mathbf{X}^\top \mathbf{X}_k$ and $\Lambda_0^* := \mathbf{P}_\mathbf{X} \Lambda_0$,

$$\begin{aligned}
& \frac{1}{n^3 T^3} \mathbb{E} \left[\left(\sum_{i=1}^n \text{tr}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top - \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \mathbf{A}_{ii}(r_1, r_2) \right)^2 \right] \\
&= \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n \text{tr}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top - \boldsymbol{\Sigma}_{\mathcal{D}, ii})^2 \mathbf{A}_{ii}^2(r_1, r_2) \right] \\
&= \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n \boldsymbol{\varepsilon}_i^\top \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \boldsymbol{\varepsilon}_i \mathbf{A}_{ii}^2(r_1, r_2) \right] - \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, ii})^2 \mathbf{A}_{ii}^2(r_1, r_2) \right] \\
&= \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T \boldsymbol{\varepsilon}_{it_1}^2 \boldsymbol{\varepsilon}_{it_2}^2 \mathbf{A}_{ii}^2(r_1, r_2) \right] - \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, ii})^2 \mathbf{A}_{ii}^2(r_1, r_2) \right] \\
&\lesssim \frac{1}{n^3 T} \mathbb{E} \left[\sum_{i=1}^n \mathbf{A}_{ii}^2(r_1, r_2) \right] \\
&= \frac{1}{n^3 T} \mathbb{E} \left[\sum_{i=1}^n (\mathbf{f}_{0, r_1}^\top \mathbf{x}_{ki}^*)^2 (\boldsymbol{\lambda}_{0, ir_2}^*)^2 \right] \\
&\leq \frac{1}{n^3 T} \mathbb{E} \left[\sum_{i=1}^n \|\mathbf{x}_{ki}^*\|_2^2 \|\mathbf{f}_{0, r_1}\|_2^2 (\boldsymbol{\lambda}_{0, ir_2}^*)^2 \right] \\
&\leq \frac{1}{n^3 T} \left(\mathbb{E} \left[\sum_{i=1}^n \|\mathbf{x}_{ki}^*\|_2^4 \|\mathbf{f}_{0, r_1}\|_2^4 \right] \right)^{\frac{1}{2}} \left(\mathbb{E} [\|\boldsymbol{\lambda}_{0, r_2}^*\|_4^4] \right)^{\frac{1}{2}} \\
&\leq \frac{1}{n^3 T} \left(\sum_{i=1}^n \left(\mathbb{E} [\|\mathbf{x}_{ki}^*\|_2^8] \right)^{\frac{1}{2}} \left(\mathbb{E} [\|\mathbf{f}_{0, r_1}\|_2^8] \right)^{\frac{1}{2}} \right)^{\frac{1}{2}} \left(\mathbb{E} [\|\boldsymbol{\lambda}_{0, r_2}^*\|_2^4] \right)^{\frac{1}{2}} \\
&= \mathcal{O} \left(\frac{1}{\sqrt{n}} \right),
\end{aligned}$$

under Assumptions **ED** and **AD(i)**. It thereby follows that $p_1 = \mathcal{O}_p(n^{-\frac{1}{4}})$. Turning now to p_2 ,

$$\begin{aligned}
& \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \sum_{i_3=1}^n \sum_{i_4 \neq i_3}^n \text{tr}(\boldsymbol{\varepsilon}_{i_1} \boldsymbol{\varepsilon}_{i_2}^\top) \text{tr}(\boldsymbol{\varepsilon}_{i_3} \boldsymbol{\varepsilon}_{i_4}^\top) \mathbf{A}_{i_1 i_2}(r_1, r_2) \mathbf{A}_{i_3 i_4}(r_1, r_2) \right] \\
&= \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, i_1 i_1} \boldsymbol{\Sigma}_{\mathcal{D}, i_2 i_2}) \mathbf{A}_{i_1 i_2}^2(r_1, r_2) \right] \\
&\quad + \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, i_1 i_1} \boldsymbol{\Sigma}_{\mathcal{D}, i_2 i_2}) \mathbf{A}_{i_1 i_2}(r_1, r_2) \mathbf{A}_{i_2 i_1}(r_1, r_2) \right] \\
&\lesssim \frac{1}{n^3 T^2} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2=1}^n \mathbf{A}_{i_1 i_2}^2(r_1, r_2) \right] + \frac{1}{n^3 T^2} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2=1}^n |\mathbf{A}_{i_1 i_2}(r_1, r_2)| |\mathbf{A}_{i_2 i_1}(r_1, r_2)| \right] \\
&\lesssim \frac{1}{n^3 T^2} \mathbb{E} [\|\mathbf{A}(r_1, r_2)\|_F^2] \\
&\leq \frac{1}{n^3 T^2} \mathbb{E} [\|\mathbf{X}_k\|_F^2 \|\mathbf{f}_{0, r_1}\|_2^2 \|\boldsymbol{\lambda}_{0, r_2}\|_2^2] \\
&\leq \frac{1}{n^3 T^2} \left(\mathbb{E} [\|\mathbf{X}_k\|_F^4] \right)^{\frac{1}{2}} \left(\mathbb{E} [\|\mathbf{f}_{0, r_1}\|_2^8] \right)^{\frac{1}{4}} \left(\mathbb{E} [\|\boldsymbol{\lambda}_{0, r_2}\|_2^8] \right)^{\frac{1}{4}} \\
&= \mathcal{O} \left(\frac{1}{n} \right),
\end{aligned}$$

under Assumptions **MD(ii)**, **ED**, and **AD(i)**. It thereby follows that $p_2 = \mathcal{O}_p(n^{-\frac{1}{2}})$ which completes the result. $\square$

Proof of Lemma C.2(ii).

$$\begin{aligned}
\frac{1}{\sqrt{nT}} \text{tr}(\mathbf{M}_{F_0} \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{B} \tilde{\mathbf{X}}_k) &= \frac{1}{\sqrt{nT}} \text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{B} \tilde{\mathbf{X}}_k \mathbf{M}_{F_0} \otimes \mathbf{P}_{\mathcal{X}}) \text{vec}(\boldsymbol{\varepsilon}) \\
&= \frac{1}{\sqrt{nT}} \text{tr} \left(\boldsymbol{\Sigma}_{\mathcal{D}} (\mathbf{B} \tilde{\mathbf{X}}_k \mathbf{M}_{F_0} \otimes \mathbf{P}_{\mathcal{X}}) \right) \\
&\quad + \frac{1}{\sqrt{nT}} \text{tr} \left(\left(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top - \boldsymbol{\Sigma}_{\mathcal{D}} \right) (\mathbf{B} \tilde{\mathbf{X}}_k \mathbf{M}_{F_0} \otimes \mathbf{P}_{\mathcal{X}}) \right). \tag{C.8}
\end{aligned}$$

It is now demonstrated that the second term in the above is $\mathcal{O}_p(n^{-\frac{1}{2}})$. Let

$$\boldsymbol{\Pi} := \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \left(\frac{1}{n} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \right)^{-1}.$$

One can write the second term on the right in (C.8) as

$$\begin{aligned}
&\sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \boldsymbol{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i_1=1}^n \sum_{i_2=1}^n (\mathbf{P}_{\mathcal{X}})_{i_1 i_2} \text{tr}((\boldsymbol{\varepsilon}_{i_1} \boldsymbol{\varepsilon}_{i_2}^\top - \boldsymbol{\Sigma}_{\mathcal{D}, i_1 i_2}) \mathbf{f}_{0, r_1} \tilde{\boldsymbol{\Lambda}}_{0, r_2}^\top \tilde{\mathbf{X}}_k \mathbf{M}_{F_0}) \right) \\
&=: \sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \boldsymbol{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i_1=1}^n \sum_{i_2=1}^n (\mathbf{P}_{\mathcal{X}})_{i_1 i_2} \text{tr}((\boldsymbol{\varepsilon}_{i_1} \boldsymbol{\varepsilon}_{i_2}^\top - \boldsymbol{\Sigma}_{\mathcal{D}, i_1 i_2}) \mathbf{A}(r_1, r_2)) \right) \\
&= \sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \boldsymbol{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i=1}^n (\mathbf{P}_{\mathcal{X}})_{ii} \text{tr}((\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top - \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \mathbf{A}(r_1, r_2)) \right) \\
&\quad + \sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \boldsymbol{\Pi}_{r_1 r_2} \left(\frac{1}{\sqrt{nT}} \frac{1}{nT} \sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n (\mathbf{P}_{\mathcal{X}})_{i_1 i_2} \text{tr}(\boldsymbol{\varepsilon}_{i_1} \boldsymbol{\varepsilon}_{i_2}^\top \mathbf{A}(r_1, r_2)) \right) \\
&=: p_1 + p_2.
\end{aligned}$$

Consider each of these terms, in turn. First, since $\|\boldsymbol{\Pi}\|_{\max} \leq \|\boldsymbol{\Pi}\|_2 = \mathcal{O}_p(1)$ under Assumption AE(ii) and AE(iii), it suffices to examine the summands (over R_0) in p_1 and p_2 for each (r_1, r_2) . Starting with p_1 ,

$$\begin{aligned}
&\frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}})_{ii}^2 \text{tr}((\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top - \boldsymbol{\Sigma}_{\mathcal{D}, ii}) \mathbf{A}(r_1, r_2))^2 \right] \\
&= \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}})_{ii}^2 \text{tr}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \mathbf{A}(r_1, r_2))^2 \right] - \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}})_{ii}^2 \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, ii} \mathbf{A}(r_1, r_2))^2 \right] \\
&\lesssim \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i=1}^n (\mathbf{P}_{\mathcal{X}})_{ii}^2 \|\mathbf{f}_{0, r_1}\|_2^2 \|\boldsymbol{\lambda}_{0, r_2}\|_2^2 \|\mathbf{X}_k\|_F^2 \right] \\
&\leq \frac{1}{n^3 T^2} \mathbb{E} [\|\mathbf{f}_{0, r_1}\|_2^2 \|\boldsymbol{\lambda}_{0, r_2}\|_2^2 \|\mathbf{X}_k\|_F^2] \\
&\leq \frac{1}{n^3 T^2} (\mathbb{E} [\|\mathbf{f}_{0, r_1}\|_2^8])^{\frac{1}{4}} (\mathbb{E} [\|\boldsymbol{\lambda}_{0, r_2}\|_2^8])^{\frac{1}{4}} (\mathbb{E} [\|\mathbf{X}_k\|_F^4])^{\frac{1}{2}} \\
&= \mathcal{O} \left(\frac{1}{n} \right),
\end{aligned}$$

under Assumptions **MD(ii)**, **ED**, and **AD(i)**. It thereby follows that $p_1 = \mathcal{O}_p(n^{-\frac{1}{2}})$. Turning now to p_2 ,

$$\begin{aligned}
& \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \sum_{i_3=1}^n \sum_{i_4 \neq i_3}^n (\mathbf{P}\mathbf{x})_{i_1 i_2} (\mathbf{P}\mathbf{x})_{i_3 i_4} \text{tr}(\boldsymbol{\varepsilon}_{i_1} \boldsymbol{\varepsilon}_{i_2}^\top \mathbf{A}(r_1, r_2)) \text{tr}(\boldsymbol{\varepsilon}_{i_3} \boldsymbol{\varepsilon}_{i_4}^\top \mathbf{A}(r_1, r_2)) \right] \\
&= \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n (\mathbf{P}\mathbf{x})_{i_1 i_2}^2 \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, i_1 i_1} \mathbf{A}^\top(r_1, r_2) \boldsymbol{\Sigma}_{\mathcal{D}, i_2 i_2} \mathbf{A}(r_1, r_2)) \right] \\
&\quad + \frac{1}{n^3 T^3} \mathbb{E} \left[\sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n (\mathbf{P}\mathbf{x})_{i_1 i_2} (\mathbf{P}\mathbf{x})_{i_2 i_1} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{D}, i_1 i_1} \mathbf{A}(r_1, r_2) \boldsymbol{\Sigma}_{\mathcal{D}, i_2 i_2} \mathbf{A}(r_1, r_2)) \right] \\
&\lesssim \frac{1}{n^3 T^3} \mathbb{E} [\|\mathbf{P}\mathbf{x}\|_F^2 \|\mathbf{A}(r_1, r_2)\|_2^2] \\
&\lesssim \frac{1}{n^3 T^2} \mathbb{E} [\|\mathbf{f}_{0, r_1}\|_2^2 \|\boldsymbol{\lambda}_{0, r_2}\|_2^2 \|\mathbf{X}_k\|_2^2] \\
&\leq \frac{1}{n^3 T^2} (\mathbb{E} [\|\mathbf{f}_{0, r_1}\|_2^8])^{\frac{1}{4}} (\mathbb{E} [\|\boldsymbol{\lambda}_{0, r_2}\|_2^8])^{\frac{1}{4}} (\mathbb{E} [\|\mathbf{X}_k\|_F^4])^{\frac{1}{2}} \\
&= \mathcal{O} \left(\frac{1}{n} \right),
\end{aligned}$$

under Assumptions **MD(ii)**, **ED**, and **AD(i)**. It thereby follows that $p_2 = \mathcal{O}_p(n^{-\frac{1}{2}})$ which completes the result. $\square$

Appendix D. Results for Section 5.1

Notation

For an $n \times m$ matrix $\mathbf{A}$, define mixed matrix norms $\|\mathbf{A}\|_{2, \infty} := \max_{1 \leq i \leq n} \|\mathbf{a}_i\|_2$ and $\|\mathbf{A}\|_{\infty, 2} := \max_{1 \leq j \leq m} \|\mathbf{a}_j\|_2$. Moreover, let $\mathbf{X}_k^* := \mathbf{M}_{\mathbf{P}\mathbf{X}\Lambda_0} \mathbf{X}_k \mathbf{M}_{F_0}$ and $\hat{\mathbf{X}}_k^* := \mathbf{Q}_{\mathbf{X}} \mathbf{M}_{\hat{\Lambda}} \mathbf{Q}_{\mathbf{X}}^\top \mathbf{X}_k \mathbf{M}_{\hat{F}}$. Note also the following inequality which is used throughout: for two matrices $\mathbf{A}$ and $\mathbf{B}$ which are $n \times m$ and $m \times T$ respectively, $\|\mathbf{AB}\|_{\max} \leq \|\mathbf{A}\|_{2, \infty} \|\mathbf{B}\|_{\infty, 2}$.⁴²

D.1. Lemmas

Lemma D.1. Under Assumptions **MD**, **ED**, **CS**, **AE**, and **AD**,

- (i) $\|\mathbf{M}_{F_0} - \mathbf{M}_{\hat{F}}\|_2 = \mathcal{O}_p(n^{-\frac{1}{2}})$;
- (ii) $\|\mathbf{M}_{\Lambda_0} - \mathbf{M}_{\hat{\Lambda}}\|_2 = \mathcal{O}_p(n^{-\frac{1}{2}})$,

where $\hat{\mathbf{F}}$ and $\hat{\Lambda}$ are defined in Section 5.1.

Lemma D.2. Under Assumptions **AD** and **IF**,

- (i) $\|\mathbf{X}_k\|_{\max}, \|\mathbf{X}_k^*\|_{\max} = \mathcal{O}_p((nT)^{\frac{1}{8}})$ for $k = 1, \dots, K$;
- (ii) $\|\mathbf{X}_k\|_{2, \infty} = \mathcal{O}_p(n^{\frac{1}{8}} \sqrt{T})$ for $k = 1, \dots, K$;
- (iii) $\|\mathbf{X}_k\|_{\infty, 2} = \mathcal{O}_p(\sqrt{nT}^{\frac{1}{8}})$ for $k = 1, \dots, K$;

⁴²Let $\mathbf{C} := \mathbf{AB}$. Then $|c_{it}| = |\mathbf{a}_i^\top \mathbf{b}_t| \leq \|\mathbf{a}_i\|_2 \|\mathbf{b}_t\|_2$, from which it then follows that $\|\mathbf{C}\|_{\max} \leq \|\mathbf{A}\|_{2, \infty} \|\mathbf{B}\|_{\infty, 2}$.

- (iv) $\|\varepsilon\|_{2,\infty} = \mathcal{O}_p(n^{\frac{1}{8}}\sqrt{T})$;
- (v) $\|\varepsilon\|_{\infty,2} = \mathcal{O}_p(\sqrt{n}T^{\frac{1}{8}})$;
- (vi) $\|\mathbf{A}_0\|_{2,\infty} = \mathcal{O}_p(n^{\frac{1}{8}})$;
- (vii) $\|\mathbf{F}_0\|_{2,\infty} = \mathcal{O}_p(T^{\frac{1}{8}})$;
- (viii) $\|\mathbf{A}_0^\top \varepsilon\|_{\infty,2} = \mathcal{O}_p(\sqrt{n}T^{\frac{1}{8}})$;
- (ix) $\|\varepsilon \mathbf{F}_0\|_{2,\infty} = \mathcal{O}_p(n^{\frac{1}{8}}\sqrt{T})$;
- (x) $\|\varepsilon\|_{\max} = \mathcal{O}_p((nT)^{\frac{1}{16}})$.

Lemma D.3. Under Assumptions **MD**, **CS**, **AE**, **AD**, and **IF**, as $n \rightarrow \infty$ with $T \geq T_{\min}$ fixed,

- (i) $\|\mathbf{e}\|_{\max} = \mathcal{O}_p(n^{\frac{1}{8}})$;
- (ii) $\|\hat{\mathbf{e}} - \mathbf{e}\|_{\max} = \mathcal{O}_p(1)$;
- (iii) $\|\hat{\mathbf{e}} - \mathbf{e}\|_F = \mathcal{O}_p(1)$,

where $\mathbf{e} := \varepsilon \mathbf{M}_{F_0}$ and $\hat{\mathbf{e}} := (\mathbf{Y} - \mathbf{X} \cdot \hat{\boldsymbol{\beta}}) \mathbf{M}_{\hat{\mathbf{F}}}$.

Lemma D.4. Under Assumptions **MD**, **CS**, **AE**, **AD**, and **IF**, as $n, T \rightarrow \infty$ and $T/n \rightarrow \gamma \in (0, \infty)$,

- (i) $\|\tilde{\mathbf{e}} - \varepsilon\|_{\max} = \mathcal{O}_p(n^{-\frac{1}{4}})$;
- (ii) $\|\hat{\mathbf{\Lambda}} \mathbf{\Gamma}^{-1} - \tilde{\mathbf{\Lambda}}_0\|_2 = \mathcal{O}_p(1)$;
- (iii) $\|\hat{\mathbf{F}} - \mathbf{F}_0 \mathbf{\Gamma}^{-\top}\|_2 = \mathcal{O}_p(1)$;
- (iv) $\|\hat{\mathbf{B}} - \mathbf{B}\|_2 = \mathcal{O}_p(n^{-\frac{3}{2}})$,

where $\mathbf{B}$ is defined in Lemma C.2, $\hat{\mathbf{B}}$ is defined analogously, and $\mathbf{\Gamma}$ is defined in the proof of D.4(ii).

D.2. Proof of Proposition 8.

First,

$$\begin{aligned} \|\hat{\mathbf{D}} - \mathbf{D}\|_2 &\leq \frac{1}{nT} \|\tilde{\mathbf{X}}^\top ((\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{F_0}) \otimes \mathbf{M}_{\hat{\mathbf{\Lambda}}}) \tilde{\mathbf{X}}\|_2 \\ &\quad + \frac{1}{nT} \|\tilde{\mathbf{X}}^\top (\mathbf{M}_{F_0} \otimes (\mathbf{M}_{\hat{\mathbf{\Lambda}}} - \mathbf{M}_{\tilde{\mathbf{\Lambda}}_0})) \tilde{\mathbf{X}}\|_2. \end{aligned} \quad (\text{D.1})$$

For the first term in the above,

$$\frac{1}{nT} \|\tilde{\mathbf{X}}^\top ((\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{F_0}) \otimes \mathbf{M}_{\hat{\mathbf{\Lambda}}}) \tilde{\mathbf{X}}\|_2 \leq \frac{1}{nT} \|\tilde{\mathbf{X}}\|_2^2 \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{F_0}\|_2.$$

Notice,

$$\|\tilde{\mathbf{X}}\|_2^2 \leq \|\tilde{\mathbf{X}}\|_F^2 = \sum_{k=1}^K \|\tilde{\mathbf{X}}_k\|_F^2 = \mathcal{O}_p(nT),$$

by Lemma B.1(i). Hence, using this and Lemma D.1, it follows that the first term is $\mathcal{O}_p(1)$. Similarly for the second term in (D.1). This concludes the first part of the proposition. Turning now to the second part, decompose

$$\hat{\mathbf{V}} - \mathbf{V} = (\hat{\mathbf{V}} - \bar{\mathbf{V}}) + (\bar{\mathbf{V}} - \mathbf{V}), \quad (\text{D.2})$$

where

$$\bar{\mathbf{V}} := \frac{1}{nT} \tilde{\mathbf{X}}^\top (M_{F_0} \otimes M_{\tilde{\Lambda}_0} \mathbf{Q}_X^\top) \Gamma_T \left(\text{vec}(\mathbf{e}) \text{vec}(\mathbf{e})^\top \right) (M_{F_0} \otimes \mathbf{Q}_X M_{\tilde{\Lambda}_0}) \tilde{\mathbf{X}}.$$

To establish that the first term in (D.2) is $\mathcal{O}_p(1)$, it is sufficient to show that $\hat{V}_{k_1 k_2} - \bar{V}_{k_1 k_2} = \mathcal{O}_p(1)$ for $k_1, k_2 = 1, \dots, K$.

$$\begin{aligned} \hat{V}_{k_1 k_2} - \bar{V}_{k_1 k_2} &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} \hat{e}_{it_2} - e_{it_1} e_{it_2}) \hat{x}_{k_1 it_1}^* \hat{x}_{k_2 it_2}^* \\ &\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T e_{it_1} e_{it_2} (\hat{x}_{k_1 it_1}^* \hat{x}_{k_2 it_2}^* - x_{k_1 it_1} x_{k_2 it_2}) \\ &\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} \hat{e}_{it_2} - e_{it_1} e_{it_2}) (\hat{x}_{k_1 it_1}^* \hat{x}_{k_2 it_2}^* - x_{k_1 it_1}^* x_{k_2 it_2}^*) \\ &=: p_1 + p_2 + p_3. \end{aligned} \tag{D.3}$$

Each of these terms is now examined. First,

$$\begin{aligned} p_1 &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} \hat{e}_{it_2} - e_{it_1} e_{it_2}) x_{k_1 it_1}^* x_{k_2 it_2}^* \\ &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T ((\hat{e}_{it_1} - e_{it_1})(\hat{e}_{it_2} - e_{it_2}) + (\hat{e}_{it_1} - e_{it_1})e_{it_2} + e_{it_1}(\hat{e}_{it_2} - e_{it_2})) x_{k_1 it_1}^* x_{k_2 it_2}^* \\ &=: p_1^{(1)} + p_1^{(2)} + p_1^{(3)}. \end{aligned}$$

Consider the first term.

$$\begin{aligned} |p_1^{(1)}| &\leq \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T |\hat{e}_{it_1} - e_{it_1}| |\hat{e}_{it_2} - e_{it_2}| |x_{k_1 it_1}^*| |x_{k_2 it_2}^*| \\ &\leq \frac{1}{n} \|\mathbf{X}_{k_1}^*\|_{\max} \|\mathbf{X}_{k_2}^*\|_{\max} \|\hat{\mathbf{e}} - \mathbf{e}\|_F^2 \\ &= \mathcal{O}_p(1), \end{aligned}$$

using Lemmas D.2(i) and D.3(iii). Consider the second term. First notice,

$$\begin{aligned} p_1^{(2)} &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) e_{it_2} x_{k_1 it_1}^* x_{k_2 it_2}^* \\ &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) \varepsilon_{it_2} x_{k_1 it_1}^* x_{k_2 it_2}^*. \end{aligned}$$

Now,

$$\begin{aligned} |p_1^{(2)}| &\leq \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T |\hat{e}_{it_1} - e_{it_1}| |\varepsilon_{it_2}| |x_{k_1 it_1}^*| |x_{k_2 it_2}^*| \\ &\leq \frac{1}{nT} \|\hat{\mathbf{e}} - \mathbf{e}\|_{\max} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T |\varepsilon_{it_2}| |x_{k_1 it_1}^*| |x_{k_2 it_2}^*| \\ &= \mathcal{O}_p(1), \end{aligned}$$

which follows using Lemma D.3(ii), and because

$$\mathbb{E} \left[\sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T |\varepsilon_{it_2}| |x_{k_1 it_1}^*| |x_{k_2 it_2}^*| \right] = \mathcal{O}(n),$$

using Assumptions IF(i) and IF(iii). Similarly for $p_1^{(3)}$ from which it then follows that $p_1 = \mathcal{O}_p(1)$. Next,

$$\begin{aligned} p_2 &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T e_{it_1} e_{it_2} (\hat{x}_{k_1 it_1}^* \hat{x}_{k_2 it_2}^* - x_{k_1 it_1}^* x_{k_2 it_2}^*) \\ &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T e_{it_1} e_{it_2} ((\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*)(\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) + (\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*) x_{k_2 it_2}^*) \\ &\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T e_{it_1} e_{it_2} x_{k_1 it_1}^* (\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) \\ &=: p_2^{(1)} + p_2^{(2)} + p_2^{(3)}. \end{aligned}$$

Consider the first term.

$$\begin{aligned} |p_2^{(1)}| &\leq \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T |e_{it_1}| |e_{it_2}| |\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*| |\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*| \\ &\leq \frac{1}{n} \|e\|_{\max}^2 \|\hat{\mathbf{X}}_{k_1}^* - \mathbf{X}_{k_1}^*\|_F \|\hat{\mathbf{X}}_{k_2}^* - \mathbf{X}_{k_2}^*\|_F \\ &= \mathcal{O}_p(1), \end{aligned}$$

using Lemma D.3(i) and

$$\|\hat{\mathbf{X}}_k^* - \mathbf{X}_k^*\|_F \leq \|\mathbf{M}_{\hat{\Lambda}} - \mathbf{M}_{\hat{\Lambda}_0}\|_2 \|\mathbf{X}_k\|_F + \|\mathbf{X}_k\|_F \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_F = \mathcal{O}_p(1), \quad (\text{D.4})$$

using Lemmas B.1(i) and D.1. Consider the second term.

$$\begin{aligned} |p_2^{(2)}| &\leq \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T |e_{it_1}| |e_{it_2}| |\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*| |x_{k_2 it_2}^*| \\ &\leq \frac{1}{n} \|e\|_{\max}^2 \|\hat{\mathbf{X}}_{k_1}^* - \mathbf{X}_{k_1}^*\|_F \|\mathbf{X}_{k_2}^*\|_F \\ &= \mathcal{O}_p(1), \end{aligned}$$

using Lemmas B.1(i), D.3(i), and (D.4). Similarly for $p_2^{(3)}$ which establishes that $p_2 = \mathcal{O}_p(1)$. Finally,

$$\begin{aligned}
p_3 &= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} \hat{e}_{it_2} - e_{it_1} e_{it_2}) (\hat{x}_{k_1 it_1}^* \hat{x}_{k_1 it_2}^* - x_{k_2 it_1}^* x_{k_2 it_2}^*) \\
&= \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) (\hat{e}_{it_2} - e_{it_2}) (\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*) (\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) (\hat{e}_{it_2} - e_{it_2}) (\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*) x_{k_2 it_2}^* \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) (\hat{e}_{it_2} - e_{it_2}) x_{k_1 it_1}^* (\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) e_{it_2} (\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*) (\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) e_{it_2} (\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*) x_{k_2 it_2}^* \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T (\hat{e}_{it_1} - e_{it_1}) e_{it_2} x_{k_1 it_1}^* (\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T e_{it_1} (\hat{e}_{it_2} - e_{it_2}) (\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*) (\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T e_{it_1} (\hat{e}_{it_2} - e_{it_2}) (\hat{x}_{k_1 it_1}^* - x_{k_1 it_1}^*) x_{k_2 it_2}^* \\
&\quad + \frac{1}{nT} \sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T e_{it_1} (\hat{e}_{it_2} - e_{it_2}) x_{k_1 it_1}^* (\hat{x}_{k_2 it_2}^* - x_{k_2 it_2}^*) \\
&=: p_3^{(1)} + \dots + p_3^{(9)}.
\end{aligned}$$

Consider the first term.

$$|p_3^{(1)}| \leq \frac{1}{n} \|\hat{\mathbf{e}} - \mathbf{e}\|_{\max}^2 \|\hat{\mathbf{X}}_{k_1}^* - \mathbf{X}_{k_1}^*\|_F \|\hat{\mathbf{X}}_{k_2}^* - \mathbf{X}_{k_2}^*\|_F = \mathcal{O}_p(1),$$

using (D.4) and Lemma D.3(ii). For term $p_3^{(2)}$,

$$|p_3^{(2)}| \leq \frac{1}{n} \|\hat{\mathbf{e}} - \mathbf{e}\|_{\max}^2 \|\hat{\mathbf{X}}_{k_1}^* - \mathbf{X}_{k_1}^*\|_F \|\mathbf{X}_{k_2}^*\|_F = \mathcal{O}_p(1),$$

using (D.4) and Lemmas B.1(i) and D.3(ii). Similar steps establish term $p_3^{(3)}, \dots, p_3^{(9)}$ are also $\mathcal{O}_p(1)$, from which it then follows that $p_3 = \mathcal{O}_p(1)$, whereby $\|\hat{\mathbf{V}} - \bar{\mathbf{V}}\|_2 = \mathcal{O}_p(1)$. Turning now to the second term in (D.2), $\|\bar{\mathbf{V}} - \mathbf{V}\|_2 = \mathcal{O}_p(1)$ using similar steps to (B.22) in the proof of Theorem 1. This completes the result.

D.3. Proof of Proposition 9.

Let

$$\begin{aligned}
\mathbf{A}_k &:= \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\mathbf{A}}_0} \tilde{\mathbf{X}}_k \mathbf{B} \mathbf{Q}_{\mathcal{X}}^\top \\
\hat{\mathbf{A}}_k &:= \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\hat{\mathbf{A}}} \tilde{\mathbf{X}}_k \hat{\mathbf{B}} \mathbf{Q}_{\mathcal{X}}^\top \\
\mathbf{B} &:= \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\mathbf{A}}_0^\top \tilde{\mathbf{A}}_0)^{-1} \tilde{\mathbf{A}}_0^\top \\
\hat{\mathbf{B}} &:= \hat{\mathbf{F}} (\hat{\mathbf{F}}^\top \hat{\mathbf{F}})^{-1} (\hat{\mathbf{A}}^\top \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^\top \\
\mathbf{C}_k &:= (\mathbf{I}_T \otimes \mathbf{A}_k) \\
\hat{\mathbf{C}}_k &:= (\mathbf{I}_T \otimes \hat{\mathbf{A}}_k).
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\psi}_k^{(1)} - \psi_k^{(1)} &= \frac{1}{\sqrt{nT}} \text{tr} \left(\Gamma_{b_T} (\text{vec}(\check{\boldsymbol{\varepsilon}}) \text{vec}(\check{\boldsymbol{\varepsilon}})^\top) \hat{\mathbf{C}}_k \right) - \frac{1}{\sqrt{nT}} \text{tr} (\boldsymbol{\Sigma}_{\mathcal{D}} \mathbf{C}_k) \\
&= \frac{1}{\sqrt{nT}} \text{tr} \left(\left(\Gamma_{b_T} (\text{vec}(\check{\boldsymbol{\varepsilon}}) \text{vec}(\check{\boldsymbol{\varepsilon}})^\top) - \Gamma_{b_T} (\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) \right) (\hat{\mathbf{C}}_k - \mathbf{C}_k) \right) \\
&\quad + \frac{1}{\sqrt{nT}} \text{tr} \left(\left(\Gamma_{b_T} (\text{vec}(\check{\boldsymbol{\varepsilon}}) \text{vec}(\check{\boldsymbol{\varepsilon}})^\top) - \Gamma_{b_T} (\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) \right) \mathbf{C}_k \right) \\
&\quad + \frac{1}{\sqrt{nT}} \text{tr} \left(\Gamma_{b_T} (\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) (\hat{\mathbf{C}}_k - \mathbf{C}_k) \right) \\
&\quad + \frac{1}{\sqrt{nT}} \text{tr} \left(\left(\Gamma_{b_T} (\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) - \boldsymbol{\Sigma}_{\mathcal{D}} \right) \mathbf{C}_k \right) \\
&=: q_{k,1} + q_{k,2} + q_{k,3} + q_{k,4}.
\end{aligned}$$

Now, notice that

$$\|\Gamma_{b_T}(\mathbf{a}\mathbf{b}^\top)\|_2 = \|\text{diag}(\mathbf{a})(\boldsymbol{\Omega} \otimes \mathbf{I}_n)\text{diag}(\mathbf{b})\|_2 \leq \|\mathbf{a}\|_\infty \|\boldsymbol{\Omega}\|_2 \|\mathbf{b}\|_\infty \leq 2b_T \|\mathbf{a}\|_\infty \|\mathbf{b}\|_\infty,$$

which follows using $\|\boldsymbol{\Omega}\|_2 \leq 2b_T$.⁴³ This latter property can be established by noticing that, since $\boldsymbol{\Omega}$ is symmetric and non-negative, its largest eigenvalue is bounded by its largest row sum by the Gershgorin Circle Theorem. Hence, $\|\boldsymbol{\Omega}\|_\infty = 2b_T - 1 \leq 2b_T$. Using this one can establish

$$\|\Gamma_{b_T}(\text{vec}(\check{\boldsymbol{\varepsilon}}) \text{vec}(\check{\boldsymbol{\varepsilon}})^\top) - \Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top)\|_2 = \mathcal{O}_p(n^{-\frac{1}{8}} b_T) = \mathcal{O}_p(1), \quad (\text{D.5})$$

as $n, T, b_T \rightarrow \infty$ and $T/n \rightarrow \gamma \in (0, \infty)$, $b_T^8/n \rightarrow 0$, which follows applying Lemmas D.2(x) and D.4(i), and under Assumption IF(iv). Using this,

$$\begin{aligned}
|q_{k,1}| &\lesssim \|\Gamma_{b_T}(\text{vec}(\check{\boldsymbol{\varepsilon}}) \text{vec}(\check{\boldsymbol{\varepsilon}})^\top) - \Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top)\|_2 \|\hat{\mathbf{C}}_k - \mathbf{C}_k\|_2 = \mathcal{O}_p(1) \\
|q_{k,2}| &\lesssim \|\Gamma_{b_T}(\text{vec}(\check{\boldsymbol{\varepsilon}}) \text{vec}(\check{\boldsymbol{\varepsilon}})^\top) - \Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top)\|_2 \|\mathbf{C}_k\|_2 = \mathcal{O}_p(1) \\
|q_{k,3}| &\lesssim \|\Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top)\|_2 \|\hat{\mathbf{C}}_k - \mathbf{C}_k\|_2 = \mathcal{O}_p(1),
\end{aligned}$$

⁴³Where, for an $n \times 1$ vector $\mathbf{a}$, $\text{diag}(\mathbf{a})$ is an $n \times n$ diagonal matrix containing the entries of $\mathbf{a}$ along its diagonal.

which follows because

$$\begin{aligned}
\|\hat{\mathbf{C}}_k - \mathbf{C}_k\|_2 &\leq \|(\mathbf{I}_T \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\hat{\Lambda}} \tilde{\mathbf{X}}_k (\hat{\mathbf{B}} - \mathbf{B}) \mathbf{Q}_{\mathcal{X}}^\top)\|_2 + \|(\mathbf{I}_T \otimes \mathbf{Q}_{\mathcal{X}} (\mathbf{M}_{\hat{\Lambda}} - \mathbf{M}_{\tilde{\Lambda}_0}) \tilde{\mathbf{X}}_k \mathbf{B} \mathbf{Q}_{\mathcal{X}}^\top)\|_2 \\
&\leq \|\tilde{\mathbf{X}}_k\|_2 \|\hat{\mathbf{B}} - \mathbf{B}\|_2 + \|\tilde{\mathbf{X}}_k\|_2 \|\mathbf{M}_{\hat{\Lambda}} - \mathbf{M}_{\tilde{\Lambda}_0}\|_2 \|\mathbf{B}\|_2 \\
&= \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

using Lemmas B.1(i), D.1(ii), and D.4(iv), as well as

$$\|\mathbf{C}_k\|_2 \lesssim \frac{1}{nT} \|\tilde{\mathbf{X}}_k\|_2 \|\mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\tilde{\mathbf{\Lambda}}_0\|_2 = \mathcal{O}_p(1), \quad (\text{D.6})$$

by Lemmas B.1(i) and B.1(ii), and Assumptions AE(ii) and AE(iii). For $q_{k,4}$,

$$q_{k,4} = \frac{1}{\sqrt{nT}} \text{tr} \left(\left(\Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) - \Gamma_{b_T}(\boldsymbol{\Sigma}_{\mathcal{D}}) \right) \mathbf{C}_k \right) + \frac{1}{\sqrt{nT}} \text{tr} \left((\Gamma_{b_T}(\boldsymbol{\Sigma}_{\mathcal{D}}) - \boldsymbol{\Sigma}_{\mathcal{D}}) \mathbf{C}_k \right).$$

Using similar steps to those in the proof of Lemma C.2(i), the first term can be shown to be $\mathcal{O}_p(1)$ as $n, T, b_T \rightarrow \infty$ and $T/n \rightarrow \gamma \in (0, \infty)$, $b_T^8/n \rightarrow 0$. For the second term,

$$\frac{1}{\sqrt{nT}} |\text{tr}((\Gamma_{b_T}(\boldsymbol{\Sigma}_{\mathcal{D}}) - \boldsymbol{\Sigma}_{\mathcal{D}}) \mathbf{C}_k)| \leq \|\Gamma_{b_T}(\boldsymbol{\Sigma}_{\mathcal{D}}) - \boldsymbol{\Sigma}_{\mathcal{D}}\|_2 \|\mathbf{C}_k\|_2 = \mathcal{O}_p(1),$$

as $n, T, b_T \rightarrow \infty$ and $T/n \rightarrow \gamma \in (0, \infty)$, $b_T^8/n \rightarrow 0$, using (D.6) and Assumption IF(iv). Similar steps establish that $|\hat{\psi}_k^{(2)} - \psi_k^{(2)}| = \mathcal{O}_p(1)$. This concludes the first and second part of the proposition. For the third part,

$$\begin{aligned}
\hat{\mathbf{D}} - \mathbf{D} &= \frac{1}{nT} \tilde{\mathbf{X}}^\top ((\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}) \otimes \mathbf{M}_{\hat{\Lambda}}) \tilde{\mathbf{X}} + \frac{1}{nT} \tilde{\mathbf{X}}^\top (\mathbf{M}_{\mathbf{F}_0} \otimes (\mathbf{M}_{\hat{\Lambda}} - \mathbf{M}_{\tilde{\Lambda}_0})) \tilde{\mathbf{X}} \\
&=: \mathbf{P}_1 + \mathbf{P}_2.
\end{aligned}$$

Consider each of these terms.

$$\|\mathbf{P}_1\|_2 \leq \frac{1}{nT} \|\tilde{\mathbf{X}}\|_2^2 \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_2 \leq \frac{1}{nT} \left(\sum_{k=1}^K \|\tilde{\mathbf{X}}_k\|_F^2 \right) \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_2 = \mathcal{O}_p(1),$$

using Lemmas B.1(i) and D.1(i). Similarly for $\mathbf{P}_2$. For the last part of the proposition it is sufficient to show that $\hat{V}_{k_1 k_2} - V_{k_1 k_2} = \mathcal{O}_p(1)$ for $k_1, k_2 = 1, \dots, K$.

$$\begin{aligned}
\hat{V}_{k_1 k_2} - V_{k_1 k_2} &= \frac{1}{nT} \text{vec}(\hat{\mathbf{X}}_{k_1}^*)^\top \left(\Gamma_{b_T}(\text{vec}(\tilde{\boldsymbol{\varepsilon}}) \text{vec}(\tilde{\boldsymbol{\varepsilon}})^\top) - \Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) \right) \text{vec}(\hat{\mathbf{X}}_{k_2}^*) \\
&\quad + \frac{1}{nT} \text{vec}(\hat{\mathbf{X}}_{k_1}^*)^\top \Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) \text{vec}(\hat{\mathbf{X}}_{k_2}^* - \mathbf{X}_{k_2}^*) \\
&\quad + \frac{1}{nT} \text{vec}(\hat{\mathbf{X}}_{k_1}^* - \mathbf{X}_{k_1}^*)^\top \Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) \text{vec}(\mathbf{X}_{k_2}^*) \\
&\quad + \left(\frac{1}{nT} \text{vec}(\mathbf{X}_{k_1}^*)^\top \Gamma_{b_T}(\text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top) \text{vec}(\mathbf{X}_{k_2}^*) - V_{k_1 k_2} \right) \\
&=: r_1 + r_2 + r_3 + r_4.
\end{aligned}$$

Consider each of these terms.

$$\begin{aligned}
|r_1| &\leq \frac{1}{nT} \|\Gamma_{b_T}(\text{vec}(\check{\epsilon})\text{vec}(\check{\epsilon})^\top) - \Gamma_{b_T}(\text{vec}(\epsilon)\text{vec}(\epsilon)^\top)\|_2 \|\hat{\mathbf{X}}_{k_1}^*\|_F \|\hat{\mathbf{X}}_{k_2}^*\|_F = \mathcal{O}_p(1), \\
|r_2| &\leq \frac{1}{nT} \|\Gamma_{b_T}(\text{vec}(\epsilon)\text{vec}(\epsilon)^\top)\|_2 \|\hat{\mathbf{X}}_{k_1}^*\|_F \|\hat{\mathbf{X}}_{k_2}^* - \mathbf{X}_{k_2}^*\|_F = \mathcal{O}_p(1), \\
|r_3| &\leq \frac{1}{nT} \|\Gamma_{b_T}(\text{vec}(\epsilon)\text{vec}(\epsilon)^\top)\|_2 \|\hat{\mathbf{X}}_{k_1}^* - \mathbf{X}_{k_1}^*\|_F \|\mathbf{X}_{k_2}^*\|_F = \mathcal{O}_p(1),
\end{aligned}$$

using $\|\hat{\mathbf{X}}_k^*\|_F \leq \|\mathbf{X}_k\|_F$,

$$\|\hat{\mathbf{X}}_k^* - \mathbf{X}_k^*\|_F \leq \|\mathbf{M}_{\hat{\Lambda}} - \mathbf{M}_{\tilde{\Lambda}_0}\|_2 \|\mathbf{X}_k\|_F + \|\mathbf{X}_k\|_F \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_2,$$

Lemmas B.1(i), D.1, D.2(x) and D.4(i), as well as (D.5). For r_4 ,

$$\begin{aligned}
r_4 &= \frac{1}{nT} \text{vec}(\mathbf{X}_{k_1}^*)^\top (\Gamma_{b_T}(\text{vec}(\epsilon)\text{vec}(\epsilon)^\top) - \Gamma_{b_T}(\Sigma_{\mathcal{D}})) \text{vec}(\mathbf{X}_{k_2}^*) \\
&\quad + \frac{1}{nT} \text{vec}(\mathbf{X}_{k_1}^*)^\top (\Gamma_{b_T}(\Sigma_{\mathcal{D}}) - \Sigma_{\mathcal{D}}) \text{vec}(\mathbf{X}_{k_2}^*). \tag{D.7}
\end{aligned}$$

The first term in the above can be shown to be $\mathcal{O}_p(1)$ similar to (B.22) in the proof of Theorem 1. For the second term

$$\frac{1}{nT} |\text{vec}(\mathbf{X}_{k_1}^*)^\top (\Gamma_{b_T}(\Sigma_{\mathcal{D}}) - \Sigma_{\mathcal{D}}) \text{vec}(\mathbf{X}_{k_2}^*)| \leq \frac{1}{nT} \|\mathbf{X}_{k_1}^*\|_F \|\mathbf{X}_{k_2}^*\|_F \|\Gamma_{b_T}(\Sigma_{\mathcal{D}}) - \Sigma_{\mathcal{D}}\|_2 = \mathcal{O}_p(1),$$

using Lemmas B.1(i) and Assumption IF(iv). This completes the result. $\square$

D.4. Proofs of Lemmas D.1 - D.4

Proof of Lemma D.1. Similar to Proposition 2, under Assumptions MD, ED, CS, and AE, Lemma S.1 in Moon and Weidner (2017) is applicable which provides the following expansion of $\mathbf{M}_{\hat{\mathbf{F}}}$:

$$\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0} = \mathbf{B}_0^{(1)} + \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{B}_k^{(1)} + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),$$

where

$$\mathbf{B}_k^{(1)} := -\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{X}}_k^\top \tilde{\Lambda}_0 (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top - \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{X}}_k \mathbf{M}_{\mathbf{F}_0},$$

for $k = 0, \dots, K$ with $\tilde{\mathbf{X}}_0 := \tilde{\epsilon}$. Using Lemmas B.1(i) and B.1(ii), as well as Assumptions AE(ii), and AE(iii), $\|\mathbf{B}_k^{(1)}\|_2 = \mathcal{O}_p(1)$ for $k = 1, \dots, K$. Moreover, $\|\mathbf{B}_0^{(1)}\|_2 = \mathcal{O}_p(n^{-\frac{1}{2}})$ using Lemmas B.1(ii), C.1(ii), C.1(iv), and C.1(vi), as well as Assumptions AE(ii) and AE(iii). Thus, applying Theorem 1,

$$\|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right).$$

Similarly, one can obtain the symmetric result $\|\mathbf{M}_{\hat{\Lambda}} - \mathbf{M}_{\tilde{\Lambda}_0}\|_2 = \mathcal{O}_p(n^{-\frac{1}{2}})$. $\square$

Proof of Lemma D.2(i).

$$\begin{aligned}\mathbb{E} [\|\mathbf{X}_k\|_{\max}] &\leq \left(\mathbb{E} \left[\max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |x_{kit}|^8 \right] \right)^{\frac{1}{8}} \\ &\leq \left(\mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T |x_{kit}|^8 \right] \right)^{\frac{1}{8}} \\ &\lesssim (nT)^{\frac{1}{8}},\end{aligned}$$

which follows using Jensen's inequality and Assumption IF(i). The result then follows by Markov's inequality. Similarly for $\mathbf{X}_k^*$. $\square$

Proof of Lemma D.2(ii).

$$\begin{aligned}\mathbb{E} [\|\mathbf{X}_k\|_{2,\infty}] &\leq \left(\mathbb{E} \left[\max_{1 \leq i \leq n} \|\mathbf{x}_{ki}\|_2^8 \right] \right)^{\frac{1}{8}} \\ &\leq \left(\mathbb{E} \left[\sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T \sum_{t_3=1}^T \sum_{t_4=1}^T x_{kit_1}^2 x_{kit_2}^2 x_{kit_3}^2 x_{kit_4}^2 \right] \right)^{\frac{1}{8}} \\ &\lesssim n^{\frac{1}{8}} \sqrt{T},\end{aligned}$$

under Assumption IF(i). $\square$

Proof of Lemma D.2(iii). Analogous to the proof of Lemma D.2(ii). $\square$

Proof of Lemma D.2(iv).

$$\begin{aligned}\mathbb{E} [\|\boldsymbol{\varepsilon}\|_{2,\infty}] &\leq \left(\mathbb{E} \left[\max_{1 \leq i \leq n} \|\boldsymbol{\varepsilon}_i\|_2^8 \right] \right)^{\frac{1}{8}} \\ &\leq \left(\mathbb{E} \left[\sum_{i=1}^n \sum_{t_1=1}^T \sum_{t_2=1}^T \sum_{t_3=1}^T \sum_{t_4=1}^T \varepsilon_{it_1}^2 \varepsilon_{it_2}^2 \varepsilon_{it_3}^2 \varepsilon_{it_4}^2 \right] \right)^{\frac{1}{8}} \\ &\lesssim n^{\frac{1}{8}} \sqrt{T},\end{aligned}$$

which follows using Jensen's inequality and Assumption IF(iii). The result then follows by Markov's inequality. $\square$

Proof of Lemma D.2(v). Analogous to the proof of Lemma D.2(iv). $\square$

Proof of Lemma D.2(vi).

$$\begin{aligned}\mathbb{E} [\|\mathbf{\Lambda}_0\|_{2,\infty}] &\leq \left(\mathbb{E} \left[\max_{1 \leq i \leq n} \|\boldsymbol{\lambda}_{0,i}\|_2^8 \right] \right)^{\frac{1}{8}} \\ &\leq \left(\mathbb{E} \left[\sum_{i=1}^n \sum_{r_1=1}^{R_0} \sum_{r_2=1}^{R_0} \sum_{r_3=1}^{R_0} \sum_{r_4=1}^{R_0} \lambda_{0,ir_1}^2 \lambda_{0,ir_2}^2 \lambda_{0,ir_3}^2 \lambda_{0,ir_4}^2 \right] \right)^{\frac{1}{8}} \\ &\lesssim n^{\frac{1}{8}},\end{aligned}$$

which follows using Jensen's inequality and Assumption AD(i). The result then follows by Markov's inequality. $\square$

Proof of Lemma D.2(vii). Analogous to the proof of Lemma D.2(vi). $\square$

Proof of Lemma D.2(viii). First it is demonstrated that under Assumption IF(iii)

$$\sup_{\|\mathbf{v}\|_2=1} \mathbb{E}[(\mathbf{v}^\top \boldsymbol{\varepsilon}_t)^8 | \mathcal{D}_{nT}] \quad (\text{D.8})$$

is uniformly bounded. Using a Rosenthal inequality (see, e.g., 9.7.c in Lin (2010))

$$\mathbb{E} \left[\left| \sum_{i=1}^n v_i \varepsilon_{it} \right|^8 | \mathcal{D}_{nT} \right] \lesssim \left(\sum_{i=1}^n \mathbb{E}[v_i^2 \varepsilon_{it}^2 | \mathcal{D}_{nT}] \right)^4 + \sum_{i=1}^n \mathbb{E}[|v_i \varepsilon_{it}|^8 | \mathcal{D}_{nT}].$$

Now,

$$\sum_{i=1}^n \mathbb{E}[v_i^2 \varepsilon_{it}^2 | \mathcal{D}_{nT}] = \sum_{i=1}^n v_i^2 \mathbb{E}[\varepsilon_{it}^2 | \mathcal{D}_{nT}] \leq \|\mathbf{v}\|_2^2 \|\boldsymbol{\Sigma}_{\mathcal{D}}\|_{\max} \leq c_1,$$

and

$$\sum_{i=1}^n \mathbb{E}[|v_i \varepsilon_{it}|^8 | \mathcal{D}_{nT}] \leq \sum_{i=1}^n v_i^8 \mathbb{E}[\varepsilon_{it}^8 | \mathcal{D}_{nT}] \leq \|\mathbf{v}\|_2^2 \left(\max_{1 \leq i \leq n} \max_{1 \leq t \leq T} \xi_{\mathcal{D},it}^8 \right) \leq c_2,$$

by Assumption IF(iii). Since the choice of $\mathbf{v}$ is arbitrary (subject to being $\mathcal{D}_{nT}$ -measurable) this establishes (D.8). Using this result,

$$\begin{aligned} \mathbb{E} \left[\|\boldsymbol{\Lambda}_0^\top \boldsymbol{\varepsilon}\|_{\infty,2} \right] &\leq \left(\mathbb{E} \left[\max_{1 \leq t \leq T} \|\boldsymbol{\Lambda}_0^\top \boldsymbol{\varepsilon}_t\|_2^8 \right] \right)^{\frac{1}{8}} \\ &\leq \left(\mathbb{E} \left[\sum_{t=1}^T \|\boldsymbol{\Lambda}_0\|_2^8 \|\mathbf{A}^\top \boldsymbol{\varepsilon}_t\|_2^8 \right] \right)^{\frac{1}{8}} \\ &= \left(\mathbb{E} \left[\sum_{t=1}^T \|\boldsymbol{\Lambda}_0\|_2^8 \left(\sum_{r=1}^{R_0} (\mathbf{a}_r^\top \boldsymbol{\varepsilon}_t)^2 \right)^4 \right] \right)^{\frac{1}{8}} \\ &\lesssim \sqrt{nT}^{\frac{1}{8}}, \end{aligned}$$

using Assumption AD(i) with $\mathbf{A} := \boldsymbol{\Lambda}_0 / \|\boldsymbol{\Lambda}_0\|_2$ if $\|\boldsymbol{\Lambda}_0\|_2 > 0$, and $\mathbf{A} := \mathbf{0}$ otherwise. The result then follows by Markov's inequality. $\square$

Proof of Lemma D.2(ix). Analogous to the proof of Lemma D.2(viii). $\square$

Proof of Lemma D.2(x). Analogous to the proof of Lemma D.2(i), using Assumption IF(iii). $\square$

Proof of Lemma D.3(i).

$$\|\mathbf{e}\|_{\max} \leq \|\boldsymbol{\varepsilon}\|_{2,\infty} \|\mathbf{M}_{F_0}\|_{\infty,2} = \mathcal{O}_p(n^{\frac{1}{8}}),$$

using Lemma D.2(iv) and

$$\|\mathbf{M}_{F_0}\|_{\infty,2} \leq \max_{\mathbf{v}: \|\mathbf{v}\|_2=1} \|\mathbf{M}_{F_0} \mathbf{v}\|_2 = \|\mathbf{M}_{F_0}\|_2 = 1. \quad (\text{D.9})$$

$\square$

Proof of Lemma D.3(ii). First,

$$\begin{aligned}\hat{\mathbf{e}} - \mathbf{e} &= \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k \mathbf{M}_{\mathbf{F}_0} + \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k (\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}) \\ &\quad + \mathbf{\Lambda}_0 \mathbf{F}_0 (\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}) + \boldsymbol{\varepsilon} (\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}).\end{aligned}\tag{D.10}$$

Then

$$\begin{aligned}\|\hat{\mathbf{e}} - \mathbf{e}\|_{\max} &\leq \sum_{k=1}^K |\beta_{0,k} - \hat{\beta}_k| \|\mathbf{X}_k\|_{2,\infty} \|\mathbf{M}_{\mathbf{F}_0}\|_{\infty,2} + \sum_{k=1}^K |\beta_{0,k} - \hat{\beta}_k| \|\mathbf{X}_k\|_{2,\infty} \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_{\infty,2} \\ &\quad + \|\mathbf{\Lambda}_0\|_{2,\infty} \|\mathbf{F}_0\|_2 \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_{\infty,2} + \|\boldsymbol{\varepsilon}\|_{2,\infty} \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_{\infty,2} = \mathcal{O}_p(1),\end{aligned}$$

which follows using Theorem 1, Lemmas B.1(ii), D.2(ii), D.2(iii), D.2(iv), and D.2(vi), as well as equation (D.9) and

$$\|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_{\infty,2} \leq \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),$$

by Lemma D.1(i). $\square$

Proof of Lemma D.3(iii).

$$\begin{aligned}\|\hat{\mathbf{e}} - \mathbf{e}\|_F &\leq \sqrt{T} \|\hat{\mathbf{e}} - \mathbf{e}\|_2 \\ &= \sqrt{T} \|(\mathbf{Y} - \mathbf{X} \cdot \hat{\boldsymbol{\beta}}) \mathbf{M}_{\hat{\mathbf{F}}} - \boldsymbol{\varepsilon} \mathbf{M}_{\mathbf{F}_0}\|_2 \\ &\leq \sqrt{T} \left(\sum_{k=1}^K |\beta_{0,k} - \hat{\beta}_k| \|\mathbf{X}_k\|_2 + \|\mathbf{\Lambda}_0\|_2 \|\mathbf{F}_0\|_2 \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_2 + \|\boldsymbol{\varepsilon}\|_2 \|\mathbf{M}_{\hat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}_0}\|_2 \right) \\ &= \mathcal{O}_p(1),\end{aligned}$$

using Theorem 1, Lemmas B.1(i), B.1(ii), and D.1(i), and $\|\boldsymbol{\varepsilon}\|_2 = \mathcal{O}_p(\sqrt{nT})$ which is straightforward to show under Assumption AD(i). $\square$

Proof of Lemma D.4(i). Notice that

$$\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0 - \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 = \mathbf{\Lambda}_0^\top \mathbf{M}_{\mathbf{X}} \mathbf{\Lambda}_0 \succeq \mathbf{0},$$

by which it follows from Assumptions AE(ii) and IF(ii) that $n^{-1} \mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0 \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{\Lambda}_0}$ as $n \rightarrow \infty$, where the eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{\Lambda}_0}$ are bounded away from zero and from above by a constant. Using this, as well as Lemma B.1(i), $\|\boldsymbol{\varepsilon}\|_2 = \mathcal{O}_p(\sqrt{nT}^{\frac{1}{4}} + n^{\frac{1}{4}} \sqrt{T})$ (which is shown in the proof of Lemma B.2(i)), and Theorem 1, Lemma S.1 in Moon and Weidner (2015) can be applied to obtain the following expansion for $\mathbf{M}_{\hat{\mathbf{F}}}$

$$\mathbf{M}_{\hat{\mathbf{F}}} = \mathbf{M}_{\mathbf{F}_0} + \mathbf{B}^{(1)} + \mathbf{B}^{(2)} + \mathbf{B}^{(3)} + \mathcal{O}_p\left(\frac{1}{n}\right),$$

where the bound on the remainder is obtained using the result given in part (iv) of Lemma S.1. Using

$$\|\mathbf{\Lambda}_0^\top \boldsymbol{\varepsilon}\|_2 = \mathcal{O}_p(n) \quad (\text{D.11})$$

$$\|\mathbf{\Lambda}_0^\top \boldsymbol{\varepsilon} \mathbf{F}_0\|_2 = \mathcal{O}_p(n) \quad (\text{D.12})$$

$$\|\boldsymbol{\varepsilon} \mathbf{F}_0\|_2 = \mathcal{O}_p(n), \quad (\text{D.13})$$

which can be shown analogously to Lemmas C.1(ii), C.1(iv), and C.1(vi) under Assumptions MD(ii) and IF(iii), one can analyse $\mathbf{B}^{(2)}$ and $\mathbf{B}^{(3)}$ and demonstrate that $\|\mathbf{B}^{(3)}\|_2 = \mathcal{O}_p(n^{-1})$ and

$$\begin{aligned} \mathbf{B}^{(2)} &= -\mathbf{M}_{\mathbf{F}_0} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top \\ &\quad - \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} \mathbf{M}_{\mathbf{F}_0} + \mathcal{O}_p\left(\frac{1}{n}\right) \\ &=: \mathbf{B}_0^{(2)} + (\mathbf{B}_0^{(2)})^\top + \mathcal{O}_p\left(\frac{1}{n}\right). \end{aligned}$$

Moreover,

$$\mathbf{B}^{(1)} = \mathbf{B}_0^{(1)} + \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{B}_k^{(1)},$$

where

$$\mathbf{B}_k^{(1)} := -\mathbf{M}_{\mathbf{F}_0} \mathbf{X}_k^\top \mathbf{\Lambda}_0 (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top - \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} \mathbf{\Lambda}_0^\top \mathbf{X}_k \mathbf{M}_{\mathbf{F}_0},$$

for $k = 0, \dots, K$ with $\mathbf{X}_0 := \boldsymbol{\varepsilon}$. Hence, combining the above one obtains

$$\mathbf{M}_{\tilde{\mathbf{F}}} = \mathbf{M}_{\mathbf{F}_0} + \mathbf{B}_0^{(1)} + \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{B}_k^{(1)} + \mathbf{B}_0^{(2)} + (\mathbf{B}_0^{(2)})^\top + \mathcal{O}_p\left(\frac{1}{n}\right).$$

Now, using this

$$\begin{aligned} \check{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon} &= \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k + \mathbf{\Lambda}_0 \mathbf{F}_0^\top + \boldsymbol{\varepsilon} \right) \mathbf{M}_{\tilde{\mathbf{F}}} - \boldsymbol{\varepsilon} \\ &= \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k + \boldsymbol{\varepsilon} \right) \mathbf{M}_{\mathbf{F}_0} + \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k + \mathbf{\Lambda}_0 \mathbf{F}_0^\top + \boldsymbol{\varepsilon} \right) \mathbf{B}_0^{(1)} \\ &\quad + \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k + \mathbf{\Lambda}_0 \mathbf{F}_0^\top + \boldsymbol{\varepsilon} \right) \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{B}_k^{(1)} \right) - \boldsymbol{\varepsilon} \\ &\quad + \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k + \mathbf{\Lambda}_0 \mathbf{F}_0^\top + \boldsymbol{\varepsilon} \right) \left(\mathbf{B}_0^{(2)} + (\mathbf{B}_0^{(2)})^\top + \mathcal{O}_p\left(\frac{1}{n}\right) \right). \end{aligned} \quad (\text{D.14})$$

Notice that

$$\left\| \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k \right\|_{2,\infty} = \mathcal{O}_p(n^{-\frac{3}{8}}),$$

$\|\mathbf{\Lambda}_0 \mathbf{F}_0^\top\|_{2,\infty} = \mathcal{O}_p(n^{\frac{5}{8}})$, and $\|\varepsilon\|_{2,\infty} = \mathcal{O}_p(n^{\frac{5}{8}})$ using Lemmas B.1(ii), D.2(ii), D.2(iv), and D.2(vi), as well as Theorem 1. Moreover,

$$\begin{aligned} \|\mathbf{B}_0^{(2)}\|_{\infty,2} &\leq \|\mathbf{M}_{\mathbf{F}_0} \varepsilon^\top \varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\infty,2} \\ &\quad + \|\mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top \varepsilon^\top \varepsilon\|_{\infty,2} \\ &\quad + \|\mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top \varepsilon^\top \varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\infty,2} \\ &= \mathcal{O}_p\left(n^{-\frac{7}{8}}\right), \end{aligned}$$

using Lemmas B.1(ii), D.2(v), and D.2(ix), as well as Assumptions AE(ii) and AE(iii), and

$$\begin{aligned} \|\mathbf{B}_k^{(1)}\|_{\infty,2} &\leq \|\mathbf{X}_k \mathbf{\Lambda}_0 (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\infty,2} \\ &\quad + \|\mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} \mathbf{\Lambda}_0^\top \mathbf{X}_k\|_{\infty,2} \\ &\quad + \|\mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} \mathbf{\Lambda}_0^\top \mathbf{X}_k \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\infty,2} \\ &= \mathcal{O}_p\left(n^{-\frac{3}{8}}\right), \end{aligned}$$

for $k = 1, \dots, K$ using Lemmas B.1(i), B.1(ii), D.2(iii), and D.2(vii) and Assumptions AE(ii) and AE(iii). Similarly for $\|\mathbf{B}_0^{(1)}\|_{\infty,2}$. Applying this to (D.14) one obtains

$$\begin{aligned} \|\tilde{\varepsilon} - \varepsilon\|_{\max} &\leq \left\| \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k + \varepsilon \right) \mathbf{M}_{\mathbf{F}_0} - \varepsilon \right\|_{\max} + \left\| (\mathbf{\Lambda}_0 \mathbf{F}_0^\top + \varepsilon) \mathbf{B}_0^{(1)} \right\|_{\max} \\ &\quad + \left\| (\mathbf{\Lambda}_0 \mathbf{F}_0^\top + \varepsilon) \left(\sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{B}_k^{(1)} \right) \right\|_{\max} + \mathcal{O}_p\left(n^{-\frac{1}{4}}\right). \end{aligned} \quad (\text{D.15})$$

Next each of these terms is examined, in turn:

$$\left\| \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k \mathbf{M}_{\mathbf{F}_0} \right\|_{\max} \leq \left\| \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k \right\|_{\max} + \left\| \sum_{k=1}^K (\beta_{0,k} - \hat{\beta}_k) \mathbf{X}_k \mathbf{P}_{\mathbf{F}_0} \right\|_{\max} = \mathcal{O}_p(n^{-\frac{3}{8}}),$$

using Lemmas B.1(ii), D.2(i), D.2(ii), and D.2(vii), as well as Assumption AE(iii) and Theorem 1,

$$\|\varepsilon \mathbf{M}_{\mathbf{F}_0} - \varepsilon\|_{\max} = \|\varepsilon \mathbf{P}_{\mathbf{F}_0}\|_{\max} = \mathcal{O}_p(n^{-\frac{1}{4}}),$$

using Lemmas D.2(vii) and D.2(ix), as well as Assumption AE(iii),

$$\|\mathbf{\Lambda}_0 \mathbf{F}_0^\top \mathbf{B}_0^{(1)}\|_{\max} = \|\mathbf{P}_{\mathbf{\Lambda}_0} \varepsilon \mathbf{M}_{\mathbf{F}_0}\|_{\max} \leq \|\mathbf{P}_{\mathbf{\Lambda}_0} \varepsilon\|_{\max} + \|\mathbf{P}_{\mathbf{\Lambda}_0} \varepsilon \mathbf{P}_{\mathbf{F}_0}\|_{\max} = \mathcal{O}_p(n^{-\frac{1}{4}}),$$

using Lemmas D.2(vi), D.2(vii), D.2(ix), equation (D.12), as well as Assumptions AE(ii) and AE(iii),

$$\begin{aligned} \|\varepsilon \mathbf{B}_0^{(1)}\|_{\max} &\leq \|\varepsilon \varepsilon^\top \mathbf{\Lambda}_0 (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\max} \\ &\quad + \|\varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top \varepsilon^\top \mathbf{\Lambda}_0 (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\max} \\ &\quad + \|\varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} \mathbf{\Lambda}_0^\top \varepsilon\|_{\max} \\ &\quad + \|\varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\mathbf{\Lambda}_0^\top \mathbf{\Lambda}_0)^{-1} \mathbf{\Lambda}_0^\top \varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\max} \\ &= \mathcal{O}_p(n^{-\frac{1}{4}}), \end{aligned}$$

using Lemmas D.2(iv), D.2(vii), D.2(viii), D.2(ix), equation (D.12), as well as Assumptions AE(ii) and AE(iii),

$$\|\Lambda_0 \mathbf{F}_0^\top \mathbf{B}_k^{(1)}\|_{\max} = \|\mathbf{P}_{\Lambda_0} \mathbf{X}_k \mathbf{M}_{\mathbf{F}_0}\|_{\max} \leq \|\mathbf{P}_{\Lambda_0} \mathbf{X}_k\|_{\max} + \|\mathbf{P}_{\Lambda_0} \mathbf{X}_k \mathbf{P}_{\mathbf{F}_0}\|_{\max} = \mathcal{O}_p(n^{\frac{1}{4}}),$$

using Lemmas B.1(i), B.1(ii), D.2(iii), D.2(vi), D.2(vii), as well as Assumptions AE(ii) and AE(iii), and

$$\begin{aligned} \|\varepsilon \mathbf{B}_k^{(1)}\|_{\max} &\leq \|\varepsilon \mathbf{X}_k \Lambda_0 (\Lambda_0^\top \Lambda_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\max} \\ &\quad + \|\varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top \mathbf{X}_k^\top \Lambda_0 (\Lambda_0^\top \Lambda_0)^{-1} (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\max} \\ &\quad + \|\varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\Lambda_0^\top \Lambda_0)^{-1} \Lambda_0^\top \mathbf{X}_k\|_{\max} \\ &\quad + \|\varepsilon \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\Lambda_0^\top \Lambda_0)^{-1} \Lambda_0^\top \mathbf{X}_k \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} \mathbf{F}_0^\top\|_{\max} \\ &= \mathcal{O}_p(n^{\frac{1}{4}}), \end{aligned}$$

using Lemmas B.1(i), B.1(ii), D.2(ix), D.2(iii), D.2(iv), D.2(vii). Applying the above to (D.15) yields the result. $\square$

Proof of Lemma D.4(ii). Notice $\hat{\Lambda}$ satisfies

$$\frac{1}{nT} (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\beta}) (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\beta})^\top \frac{\hat{\Lambda}}{\sqrt{n}} = \frac{\hat{\Lambda}}{\sqrt{n}} \mathbf{\Pi},$$

with $\mathbf{\Pi}$ being an $R_0 \times R_0$ diagonal matrix containing the largest R_0 eigenvalues of $(nT)^{-1} (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\beta}) (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\beta})^\top$ along its diagonal. Expanding this,

$$\begin{aligned} \frac{\hat{\Lambda}}{\sqrt{n}} \mathbf{\Pi} &= \frac{1}{nT} (\tilde{\mathbf{X}} \cdot (\beta_0 - \hat{\beta}) + \tilde{\Lambda}_0 \mathbf{F}_0^\top + \tilde{\varepsilon}) (\tilde{\mathbf{X}} \cdot (\beta_0 - \hat{\beta}) + \tilde{\Lambda}_0 \mathbf{F}_0^\top + \tilde{\varepsilon})^\top \frac{\hat{\Lambda}}{\sqrt{n}} \\ &=: \mathbf{P}_1 + \dots + \mathbf{P}_8 + \frac{\tilde{\Lambda}_0}{\sqrt{n}} \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right) \left(\frac{1}{n} \tilde{\Lambda}_0^\top \hat{\Lambda} \right), \end{aligned}$$

or

$$\frac{\hat{\Lambda}}{\sqrt{n}} \mathbf{\Pi} - \frac{\tilde{\Lambda}_0}{\sqrt{n}} \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right) \left(\frac{1}{n} \tilde{\Lambda}_0^\top \hat{\Lambda} \right) = \mathbf{P}_1 + \dots + \mathbf{P}_8.$$

Now, since $\mathbf{M}_{\tilde{\Lambda}_0} \tilde{\Lambda}_0 = \mathbf{0}$,

$$\left\| \frac{1}{n} \tilde{\Lambda}_0^\top \mathbf{M}_{\hat{\Lambda}} \tilde{\Lambda}_0 \right\|_2 = \left\| \frac{1}{n} \tilde{\Lambda}_0^\top (\mathbf{M}_{\hat{\Lambda}} - \mathbf{M}_{\tilde{\Lambda}_0}) \tilde{\Lambda}_0 \right\|_2 \leq \frac{1}{n} \|\tilde{\Lambda}_0\|_2^2 \|\mathbf{P}_{\tilde{\Lambda}_0} - \mathbf{P}_{\hat{\Lambda}}\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),$$

using Lemmas B.1(ii) and D.1(ii). Hence,

$$\frac{1}{n} \tilde{\Lambda}_0^\top \tilde{\Lambda}_0 = \left(\frac{1}{n} \tilde{\Lambda}_0^\top \hat{\Lambda} \right) \left(\frac{1}{n} \tilde{\Lambda}_0^\top \hat{\Lambda} \right)^\top + \frac{1}{n} \tilde{\Lambda}_0^\top \mathbf{M}_{\hat{\Lambda}} \tilde{\Lambda}_0 = \left(\frac{1}{n} \tilde{\Lambda}_0^\top \hat{\Lambda} \right) \left(\frac{1}{n} \tilde{\Lambda}_0^\top \hat{\Lambda} \right)^\top + \mathcal{O}_p(1). \quad (\text{D.16})$$

The matrix on the left in (D.16) is invertible w.p.a.1 by Assumption AE(ii), therefore the matrix on the right must also be invertible w.p.a.1. Since $n^{-1} \tilde{\Lambda}_0^\top \hat{\Lambda}$ is asymptotically invertible, as is $T^{-1} \mathbf{F}_0^\top \mathbf{F}_0$ by

Assumption **AE(iii)**, one obtains

$$\frac{\hat{\mathbf{\Lambda}}}{\sqrt{n}}\mathbf{\Gamma}^{-1} - \frac{\tilde{\mathbf{\Lambda}}_0}{\sqrt{n}} = (\mathbf{P}_1 + \dots + \mathbf{P}_8)\bar{\mathbf{\Gamma}}^{-1}, \quad (\text{D.17})$$

with $\mathbf{\Gamma}^{-1} := \mathbf{\Pi}\bar{\mathbf{\Gamma}}^{-1}$ and

$$\bar{\mathbf{\Gamma}}^{-1} := \text{plim}_{n,T \rightarrow \infty} \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \hat{\mathbf{\Lambda}} \right)^{-1}.$$

Now,

$$\frac{1}{\sqrt{n}} \|\hat{\mathbf{\Lambda}}\mathbf{\Gamma}^{-1} - \tilde{\mathbf{\Lambda}}_0\|_2 \leq (\|\mathbf{P}_1\|_2 + \dots + \|\mathbf{P}_8\|_2) \|\bar{\mathbf{\Gamma}}^{-1}\|_2,$$

where $\|\bar{\mathbf{\Gamma}}^{-1}\|_2 = \mathcal{O}(1)$. Consider each of these terms.

$$\|\mathbf{P}_1\|_2 \leq \frac{1}{\sqrt{n}} \frac{1}{nT} \left(\sum_{k=1}^K \|\tilde{\mathbf{X}}_k\|_F^2 \right) \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2^2 \|\hat{\mathbf{\Lambda}}\|_2 = \mathcal{O}_p \left(\frac{1}{n^2} \right),$$

using Lemma **B.1(i)**, Theorem **1** and $\|\hat{\mathbf{\Lambda}}\|_2 = \sqrt{n}$.

$$\|\mathbf{P}_2\|_2 \leq \frac{1}{\sqrt{n}} \frac{1}{nT} \left(\sum_{k=1}^K \|\tilde{\mathbf{X}}_k\|_F^2 \right)^{\frac{1}{2}} \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \|\mathbf{\Lambda}_0\|_2 \|\mathbf{F}_0\|_2 \|\hat{\mathbf{\Lambda}}\|_2 = \mathcal{O}_p \left(\frac{1}{n} \right),$$

using Lemmas **B.1(i)**, **B.1(ii)**, Theorem **1** and $\|\hat{\mathbf{\Lambda}}\|_2 = \sqrt{n}$.

$$\|\mathbf{P}_3\|_2 \leq \frac{1}{\sqrt{n}} \frac{1}{nT} \left(\sum_{k=1}^K \|\tilde{\mathbf{X}}_k\|_F^2 \right)^{\frac{1}{2}} \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \|\boldsymbol{\varepsilon}\|_F \|\hat{\mathbf{\Lambda}}\|_2 = \mathcal{O}_p \left(\frac{1}{n} \right),$$

using Lemma **B.1(i)**, Theorem **1**, $\|\boldsymbol{\varepsilon}\|_F = \mathcal{O}_p(\sqrt{nT})$ and $\|\hat{\mathbf{\Lambda}}\|_2 = \sqrt{n}$.

$$\|\mathbf{P}_4\|_2 \leq \frac{1}{\sqrt{n}} \frac{1}{nT} \|\tilde{\mathbf{\Lambda}}_0\|_2 \|\tilde{\boldsymbol{\varepsilon}}\mathbf{F}_0\|_2 \|\hat{\mathbf{\Lambda}}\|_2 = \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right),$$

using Lemmas **B.1(ii)** and **C.1(vi)**, and $\|\hat{\mathbf{\Lambda}}\|_2 = \sqrt{n}$. Similarly for the remaining terms, from which the result then follows. $\square$

Proof of Lemma **D.4(iii).** Decompose

$$(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\boldsymbol{\beta}}) =: \sum_{t=1}^T s_t \mathbf{u}_t \mathbf{v}_t^\top,$$

by the singular value decomposition. Then

$$\begin{aligned} \frac{1}{n} (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\boldsymbol{\beta}})^\top \hat{\mathbf{\Lambda}} &= \frac{1}{\sqrt{n}} \left(\sum_{t=1}^T s_t \mathbf{v}_t \mathbf{u}_t^\top \mathbf{u}_1 \quad \dots \quad \sum_{t=1}^T s_t \mathbf{v}_t \mathbf{u}_t^\top \mathbf{u}_{R_0} \right) \\ &= \frac{1}{\sqrt{n}} \begin{pmatrix} s_1 \mathbf{v}_1 & \dots & s_{R_0} \mathbf{v}_{R_0} \end{pmatrix} \\ &= \hat{\mathbf{F}}. \end{aligned}$$

Thus,

$$\begin{aligned}
\hat{\mathbf{F}} &= \frac{1}{n}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} \cdot \hat{\boldsymbol{\beta}})^\top \hat{\hat{\mathbf{A}}} \\
&= \frac{1}{n}(\tilde{\mathbf{X}} \cdot (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}))^\top \hat{\hat{\mathbf{A}}} + \frac{1}{n}\mathbf{F}_0\tilde{\mathbf{A}}_0^\top \hat{\hat{\mathbf{A}}} + \frac{1}{n}\tilde{\boldsymbol{\varepsilon}}^\top \hat{\hat{\mathbf{A}}} \\
&=: \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3.
\end{aligned}$$

Now,

$$\|\mathbf{P}_1\|_2 \leq \frac{1}{n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \left(\sum_{k=1}^K \|\mathbf{X}_k\|_F^2 \right)^{\frac{1}{2}} \|\hat{\hat{\mathbf{A}}}\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),$$

using Lemma B.1(i), Theorem 1 and $\|\hat{\hat{\mathbf{A}}}\|_2 = \sqrt{n}$.

$$\mathbf{P}_2 = \frac{1}{n}\mathbf{F}_0(\hat{\hat{\mathbf{A}}}\mathbf{\Gamma}^{-1})^\top \hat{\hat{\mathbf{A}}} + \frac{1}{n}\mathbf{F}_0(\tilde{\mathbf{A}}_0 - \hat{\hat{\mathbf{A}}}\mathbf{\Gamma}^{-1})^\top \hat{\hat{\mathbf{A}}} = \mathbf{F}_0\mathbf{\Gamma}^{-\top} + \mathcal{O}_p(1),$$

using Lemmas B.1(ii) and D.4(ii), as well as $\|\hat{\hat{\mathbf{A}}}\|_2 = \sqrt{n}$. Finally,

$$\|\mathbf{P}_3\|_2 \leq \frac{1}{n}\|\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{A}}_0\|_2 + \frac{1}{n}\|\tilde{\boldsymbol{\varepsilon}}^\top (\hat{\hat{\mathbf{A}}} - \tilde{\mathbf{A}}_0\mathbf{\Gamma})\|_2 = \mathcal{O}_p(1),$$

using Lemmas C.1(ii) and D.4(ii). Combining the above gives the result. $\square$

Proof of Lemma D.4(iv).

$$\begin{aligned}
\hat{\mathbf{B}} - \mathbf{B} &= \frac{1}{nT}\hat{\hat{\mathbf{A}}} \left(\left(\frac{1}{n}\hat{\hat{\mathbf{A}}}^\top \hat{\hat{\mathbf{A}}} \right)^{-1} - \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \right) \left(\frac{1}{T}\hat{\mathbf{F}}^\top \hat{\mathbf{F}} \right)^{-1} \hat{\mathbf{F}}^\top \\
&\quad + \frac{1}{nT}\hat{\hat{\mathbf{A}}} \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \left(\left(\frac{1}{T}\hat{\mathbf{F}}^\top \hat{\mathbf{F}} \right)^{-1} - \left(\frac{1}{T}(\mathbf{F}_0\mathbf{\Gamma}^{-\top})^\top \mathbf{F}_0\mathbf{\Gamma}^{-\top} \right)^{-1} \right) \hat{\mathbf{F}}^\top \\
&\quad + \frac{1}{nT}(\hat{\hat{\mathbf{A}}} - \tilde{\mathbf{A}}_0\mathbf{\Gamma}) \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \left(\frac{1}{T}(\mathbf{F}_0\mathbf{\Gamma}^{-\top})^\top \mathbf{F}_0\mathbf{\Gamma}^{-\top} \right)^{-1} \hat{\mathbf{F}}^\top \\
&\quad + \frac{1}{nT}\tilde{\mathbf{A}}_0\mathbf{\Gamma} \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \left(\frac{1}{T}(\mathbf{F}_0\mathbf{\Gamma}^{-\top})^\top \mathbf{F}_0\mathbf{\Gamma}^{-\top} \right)^{-1} (\hat{\mathbf{F}} - \mathbf{F}_0\mathbf{\Gamma}^{-\top})^\top \\
&=: \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 + \mathbf{P}_4.
\end{aligned}$$

Consider each of these terms.

$$\begin{aligned}
\|\mathbf{P}_1\|_2 &\leq \frac{1}{nT}\|\hat{\hat{\mathbf{A}}}\|_2 \left\| \left(\frac{1}{n}\hat{\hat{\mathbf{A}}}^\top \hat{\hat{\mathbf{A}}} \right)^{-1} - \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \right\|_2 \left\| \left(\frac{1}{T}\hat{\mathbf{F}}^\top \hat{\mathbf{F}} \right)^{-1} \right\|_2 \|\hat{\mathbf{F}}\|_2 = \mathcal{O}_p\left(n^{-\frac{3}{2}}\right) \\
\|\mathbf{P}_2\|_2 &\leq \frac{1}{nT}\|\hat{\hat{\mathbf{A}}}\|_2 \left\| \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \right\|_2 \left\| \left(\frac{1}{T}\hat{\mathbf{F}}^\top \hat{\mathbf{F}} \right)^{-1} - \left(\frac{1}{T}(\mathbf{F}_0\mathbf{\Gamma}^{-\top})^\top \mathbf{F}_0\mathbf{\Gamma}^{-\top} \right)^{-1} \right\|_2 \|\hat{\mathbf{F}}\|_2 = \mathcal{O}_p\left(n^{-\frac{3}{2}}\right) \\
\|\mathbf{P}_3\|_2 &\leq \frac{1}{nT}\|\hat{\hat{\mathbf{A}}} - \tilde{\mathbf{A}}_0\mathbf{\Gamma}\|_2 \left\| \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \right\|_2 \left\| \left(\frac{1}{T}(\mathbf{F}_0\mathbf{\Gamma}^{-\top})^\top \mathbf{F}_0\mathbf{\Gamma}^{-\top} \right)^{-1} \right\|_2 \|\hat{\mathbf{F}}\|_2 = \mathcal{O}_p\left(n^{-\frac{3}{2}}\right) \\
\|\mathbf{P}_4\|_2 &\leq \frac{1}{nT}\|\tilde{\mathbf{A}}_0\|_2\|\mathbf{\Gamma}\|_2 \left\| \left(\frac{1}{n}(\tilde{\mathbf{A}}_0\mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0\mathbf{\Gamma} \right)^{-1} \right\|_2 \left\| \left(\frac{1}{T}(\mathbf{F}_0\mathbf{\Gamma}^{-\top})^\top \mathbf{F}_0\mathbf{\Gamma}^{-\top} \right)^{-1} \right\|_2 \|\hat{\mathbf{F}} - \mathbf{F}_0\mathbf{\Gamma}^{-\top}\|_2 = \mathcal{O}_p\left(n^{-\frac{3}{2}}\right),
\end{aligned}$$

using $\|\hat{\mathbf{A}}\|_2 = \sqrt{n}$, $\|\hat{\mathbf{F}}\|_2 \leq \|\hat{\mathbf{F}} - \mathbf{F}_0\mathbf{\Gamma}^{-\top}\|_2 + \|\mathbf{F}_0\|_2\|\mathbf{\Gamma}^{-1}\|_2 = \mathcal{O}_p(\sqrt{T})$, Lemmas B.1(ii), D.4(ii) and D.4(iii), Assumptions AE(ii) and AE(iii), as well as

$$\left\| \left(\frac{1}{T} \hat{\mathbf{F}}^\top \hat{\mathbf{F}} \right)^{-1} - \left(\frac{1}{T} (\mathbf{F}_0 \mathbf{\Gamma}^{-\top})^\top \mathbf{F}_0 \mathbf{\Gamma}^{-\top} \right)^{-1} \right\|_2 = \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right),$$

and

$$\left\| \left(\frac{1}{n} \hat{\mathbf{A}}^\top \hat{\mathbf{A}} \right)^{-1} - \left(\frac{1}{n} (\tilde{\mathbf{A}}_0 \mathbf{\Gamma})^\top \tilde{\mathbf{A}}_0 \mathbf{\Gamma} \right)^{-1} \right\|_2 = \mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right),$$

which are straightforward to show using Lemmas D.4(ii) and D.4(iii), as well as Assumptions AE(ii) and AE(iii). $\square$

Appendix E. Results for Section 5.3

Notation

Some of the additional notation required for the dynamic model is restated here for convenience. Recall $\theta_{0,1} := \alpha_0$, $\mathbf{Z}_1 := \mathbf{Y}_{-1} := (\mathbf{y}_0, \dots, \mathbf{y}_{T-1})$, and $\theta_{0,k+1} = \beta_{0,k}$, $\mathbf{Z}_{k+1} = \mathbf{X}_k$ for $k = 1, \dots, K$, π_J denotes $J \times 1$ vector of all zeros, except the first element, which equals to 1, and $\mathbf{W}$ is a $T \times T$ shift matrix which consists of zeros everywhere, except those elements directly above the main diagonal, which take on a value of 1, $\mathbf{G}(\alpha) := (\mathbf{I}_T - \alpha \mathbf{W})^{-1} \mathbf{W}$, $\mathbf{G} := \mathbf{G}(\alpha_0)$, $\mathbf{g}(\alpha) := (\mathbf{I}_T + \alpha \mathbf{G}(\alpha))^\top \pi_T$, $\mathbf{g} := \mathbf{g}(\alpha_0)$, $\mathbf{H}_1 := (\mathbf{X} \cdot \beta_0) \mathbf{G} + \mathbf{y}_0 \mathbf{g}^\top$, $\mathbf{H}_{k+1} := \mathbf{X}_k$ for $k = 1, \dots, K$, and $\mathbf{H}_k^* := \mathbf{Q}_X \mathbf{M}_{\tilde{\mathbf{A}}_0} \mathbf{Q}_X^\top \mathbf{H}_k \mathbf{M}_{\mathbf{F}_0}$ for $k = 1, \dots, K+1$. Moreover, define

$$\begin{aligned} \mathbf{c}_+ &:= \frac{1}{\sqrt{nT}} \tilde{\mathbf{H}}^\top (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \text{vec}(\tilde{\boldsymbol{\varepsilon}}) + \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \tilde{\boldsymbol{\varepsilon}}^\top) \pi_{K+1} \\ \Upsilon^{(1)} &:= \frac{1}{nT} \tilde{\mathbf{H}}^\top (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\mathbf{H}} \\ \Upsilon^{(2)} &:= \frac{1}{nT} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}} (\mathbf{G} \mathbf{G}^\top \otimes \mathbf{I}_{TK})) \pi_{K+1} \pi_{K+1}^\top \\ \Upsilon^{(3)} &:= \frac{1}{nT} \tilde{\mathbf{H}}^\top (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\boldsymbol{\Sigma}}_{\mathcal{E}} (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\mathbf{A}}_0}) \tilde{\mathbf{H}} \\ \Upsilon^{(4)} &:= \frac{1}{nT} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}} (\mathbf{G} \otimes \mathbf{I}_{TK}) \tilde{\boldsymbol{\Sigma}}_{\mathcal{E}} (\mathbf{G}^\top \otimes \mathbf{I}_{TK})) \pi_{K+1} \pi_{K+1}^\top, \end{aligned}$$

$\mathbf{D} := \Upsilon^{(1)}$, $\mathbf{D}_+ := \Upsilon^{(1)} + \Upsilon^{(2)}$, $\mathbf{V} := \Upsilon^{(3)}$, and $\mathbf{V}_+ := \Upsilon^{(3)} + \Upsilon^{(4)}$, with $\tilde{\mathbf{H}} := (\text{vec}(\tilde{\mathbf{H}}_1), \dots, \text{vec}(\tilde{\mathbf{H}}_{K+1}))$.

E.1. Extended Assumptions

Presented below are extensions of the main assumptions to allow for lagged outcomes as regressors.

Assumption MD*.

- (i) The parameter vector $\boldsymbol{\theta}_0$ lies in the interior of Θ_θ , where Θ_θ is a compact subset of $\mathbb{R}^{K+1}$ in which $|\alpha| < 1$.
- (ii) Assumption MD(ii).

Assumption ED*. Conditional on $\mathcal{E}_{nT}$, ε_{it} are independent over i and t , with $\mathbb{E}[\varepsilon_{it}|\mathcal{E}_{nT}] = 0$, $\mathbb{E}[\varepsilon_{it}^2|\mathcal{E}_{nT}] > 0$, and $\mathbb{E}[\varepsilon_{it}^4|\mathcal{E}_{nT}]$ uniformly bounded.

Assumption CS*.

- (i) Assumption **CS(i)**.
- (ii) $\min_{\delta \in \mathbb{R}^{K+1}: \|\delta\|_2=1} \sum_{t=T_{\min}}^T \mu_t((nT)^{-1}(\tilde{\mathbf{Z}} \cdot \delta)^\top (\tilde{\mathbf{Z}} \cdot \delta)) \geq b > 0$, w.p.a.1 as $n \rightarrow \infty$, with $T \geq T_{\min}$ fixed or $T \rightarrow \infty$.

Assumption AE*. Assumption **AE**.

Assumption AD*.

- (i) The elements of $\mathbf{H}_k$ and $\mathbf{H}_k^*$ have uniformly bounded fourth moments. Moreover, the elements of $\mathbf{Q}_X \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_X^\top \mathbf{H}_k$, $\mathbf{\Lambda}_0$, and $\mathbf{F}_0$ have uniformly bounded eighth moments.
- (ii) As $n \rightarrow \infty$, with $T \geq T_{\min}$ fixed or $T \rightarrow \infty$ and $T/n \rightarrow 0$, there exist nonstochastic matrices $\mathbb{D}$ and $\mathbb{V}$ such that $\mathbf{D} \xrightarrow{p} \mathbb{D}$ and $\mathbf{V} \xrightarrow{p} \mathbb{V}$, where the eigenvalues of $\mathbb{D}$ and $\mathbb{V}$ are bounded away from zero and from above by a constant. Otherwise, as $n, T \rightarrow \infty$ with $T/n \rightarrow \gamma \in (0, \infty)$, there exist nonstochastic matrices $\mathbb{D}_+$ and $\mathbb{V}_+$ such that $\mathbf{D}_+ \xrightarrow{p} \mathbb{D}_+$ and $\mathbf{V}_+ \xrightarrow{p} \mathbb{V}_+$, and the eigenvalues of $\mathbb{D}_+$ and $\mathbb{V}_+$ are bounded away from zero and from above by a constant.

E.2. Lemmas

Lemma E.1. Under Assumption **MD***,

- (i) $\mathbf{Y}_{-1} = \mathbf{H}_1 + \mathbf{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G} + \varepsilon \mathbf{G}$;
- (ii) $\|\mathbf{g}(\alpha)\|_2 \leq c$ and $\|\mathbf{G}(\alpha)\|_2 \leq c$.

Lemma E.2. Under Assumptions **MD***, **ED***, **CS***, and **AD***,

- (i) $(nT)^{-\frac{1}{2}} \|\tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p(n^{-\frac{1}{4}})$;
- (ii) $\|\mathbf{Z}_k\|_2, \|\tilde{\mathbf{Z}}_k\|_2, \|\mathbf{Z}_k\|_F, \|\tilde{\mathbf{Z}}_k\|_F = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K+1$;
- (iii) $\text{tr}(\tilde{\mathbf{Z}}_k^\top \tilde{\boldsymbol{\varepsilon}}) = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K+1$;
- (iv) $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = \mathcal{O}_p(1)$ as $n \rightarrow \infty$, with $T \geq T_{\min}$ fixed or $T \rightarrow \infty$;
- (v) $\|\mathbf{F}_0^\top \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{H}}_k\|_2 = \mathcal{O}_p(\sqrt{nT})$ for $k = 1, \dots, K+1$;
- (vi) $\|\tilde{\mathbf{H}}_k \tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{\Lambda}}_0\|_2 = \mathcal{O}_p(n\sqrt{T})$ for $k = 1, \dots, K+1$;
- (vii) $\|\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{F}_0\|_2 = \mathcal{O}_p(T)$.

Lemma E.3. Under Assumptions **MD***, **ED***, **CS***, **AE***, and **AD***,

- (i) $(nT)^{-\frac{1}{2}} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}) = (nT)^{-\frac{1}{2}} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\mathbf{P}_{\mathbf{F}_0} \mathbf{G} \mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0})) + \mathcal{O}_p(n^{-\frac{1}{2}})$;
- (ii) $(nT)^{-\frac{1}{2}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top) = (nT)^{-\frac{1}{2}} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{I}_{TK})) + \mathcal{O}_p(n^{-\frac{1}{2}})$;
- (iii) $(nT)^{-\frac{1}{2}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\Lambda}_0}) = \mathcal{O}_p(n^{-\frac{1}{2}})$;
- (iv) $(nT)^{-\frac{1}{2}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\Lambda}_0}) = \mathcal{O}_p((nT)^{-\frac{1}{2}})$;

- (v) $(nT)^{-1} \text{tr}(\tilde{\mathbf{H}}_k \mathbf{M}_{\mathbf{F}_0} (\tilde{\mathbf{E}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) = \mathcal{O}_p((nT)^{-\frac{1}{2}})$ for $k = 1, \dots, K+1$;
 - (vi) $(nT)^{-1} \text{tr}(\tilde{\mathbf{E}} \mathbf{G} \mathbf{M}_{\mathbf{F}_0} (\tilde{\mathbf{E}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) = (nT)^{-1} \text{tr}(\tilde{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{G}^\top \otimes \mathbf{I}_{TK})) + \mathcal{O}_p(n^{-1})$;
 - (vii) $(nT)^{-\frac{1}{2}} \text{tr}(\tilde{\mathbf{E}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{H}}_k \mathbf{B} \tilde{\mathbf{E}}) = \psi_k^{(1)} + \mathcal{O}_p(n^{-\frac{1}{4}})$ for $k = 1, \dots, K+1$;
 - (viii) $(nT)^{-\frac{1}{2}} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{E}}^\top \tilde{\mathbf{E}} \mathbf{B} \tilde{\mathbf{H}}_k) = \psi_k^{(2)} + \mathcal{O}_p(n^{-\frac{1}{2}})$ for $k = 1, \dots, K+1$,
- where $\mathbf{B} := \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top$.

Lemma E.4. Under Assumptions MD^* , ED^* , CS^* , AE^* , and AD^* ,

- (i) $\mathbf{b}_+^{(1)} = -\boldsymbol{\psi}^{(1)} + \mathcal{O}_p(n^{-\frac{1}{4}})$;
- (ii) $\mathbf{b}_+^{(2)} = \mathcal{O}_p(n^{-\frac{1}{4}})$;
- (iii) $\mathbf{b}_+^{(3)} = -\boldsymbol{\psi}^{(2)} - \boldsymbol{\xi}^{(1)} + \mathcal{O}_p(n^{-\frac{1}{4}})$.

Lemma E.5. Under Assumptions MD^* , ED^* , CS^* , AE^* , and AD^* ,

- (i) $\bar{\mathbf{c}}_+ = \mathbf{c}_+ - \boldsymbol{\xi}^{(2)} + \mathcal{O}_p(n^{-\frac{1}{2}})$;
- (ii) $\bar{\mathbf{D}}_+ = \mathbf{D}_+ + \mathcal{O}_p((nT)^{-\frac{1}{2}})$.

Lemma E.6. Under Assumptions MD^* , ED^* , CS^* , AE^* , and AD^* , $\mathbf{c}_+ \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{V}_+)$ as $n \rightarrow \infty$, with $T \geq T_{\min}$ fixed or $T \rightarrow \infty$.

E.3. Proof of Theorem 2.

First, using Lemma E.2, one can establish an expansion analogous to that obtained in Proposition 2 which reads

$$\mathcal{Q}(\boldsymbol{\theta}) = \mathcal{Q}(\boldsymbol{\theta}_0) - \frac{2}{\sqrt{nT}} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{d}_+ + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \bar{\mathbf{D}}_+ (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + r(\boldsymbol{\theta}), \quad (\text{E.1})$$

where $\mathbf{d}_+ := \bar{\mathbf{c}}_+ + \mathbf{b}_+^{(1)} + \mathbf{b}_+^{(2)} + \mathbf{b}_+^{(3)}$, and the elements of these vectors and matrices are given by

$$\begin{aligned} \bar{\mathbf{D}}_+ &:= \frac{1}{nT} \tilde{\mathbf{Z}}^\top (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \tilde{\mathbf{Z}} \\ \bar{\mathbf{c}}_+ &:= \frac{1}{\sqrt{nT}} \tilde{\mathbf{Z}}^\top (\mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \text{vec}(\tilde{\mathbf{E}}) \\ b_{+,k}^{(1)} &:= -\frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{E}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{Z}}_k \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{E}} \right) \\ b_{+,k}^{(2)} &:= -\frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{Z}}_k^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{E}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{E}} \right) \\ b_{+,k}^{(3)} &:= -\frac{1}{\sqrt{nT}} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\mathbf{E}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{E}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{Z}}_k \right), \end{aligned}$$

for $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 = \mathcal{O}_p(1)$, where $r(\boldsymbol{\theta})$ is of order $\mathcal{O}_p((nT)^{-1} (1 + \sqrt{nT} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2)^2)$, and $\tilde{\mathbf{Z}} := (\text{vec}(\tilde{\mathbf{Z}}_1), \dots, \text{vec}(\tilde{\mathbf{Z}}_{K+1}))$. Notice by Assumption $\text{CS}^*(\text{ii})$, the matrix $\bar{\mathbf{D}}_+$ is invertible and $\|\bar{\mathbf{D}}_+^{-1}\|_2 \leq c$ w.p.a.1. Define $\mathbf{a}_+ := (nT)^{-\frac{1}{2}} \bar{\mathbf{D}}_+^{-1} \mathbf{d}_+$. The stochastic order of this quantity is examined. First, it is assumed in the statement of the theorem that $\|\mathbf{c}_+\|_2 = \mathcal{O}_p(1)$. By Lemmas E.4(i), E.4(ii), and E.4(iii),

$$\mathbf{b}_+^{(1)}, \mathbf{b}_+^{(3)} = \mathcal{O}_p \left(\sqrt{\frac{T}{n}} \right),$$

and $\mathbf{b}_+^{(2)} = \mathcal{O}_p(n^{-\frac{1}{4}})$. Hence, $\|\mathbf{a}_+\|_2 = \mathcal{O}_p((nT)^{-\frac{1}{2}}) = \mathcal{O}_p(1)$. Now, using (E.1) and Lemma E.2(iv),

$$\begin{aligned}\mathcal{Q}(\hat{\boldsymbol{\theta}}) &= \mathcal{Q}(\boldsymbol{\theta}_0) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+)^{\top} \bar{\mathbf{D}}_+ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+) - \mathbf{a}_+^{\top} \bar{\mathbf{D}}_+ \mathbf{a}_+ + \mathbf{r}(\hat{\boldsymbol{\theta}}) \\ \mathcal{Q}(\boldsymbol{\theta}_0 + \mathbf{a}_+) &= \mathcal{Q}(\boldsymbol{\theta}_0) - \mathbf{a}_+^{\top} \bar{\mathbf{D}}_+ \mathbf{a}_+ + \mathbf{r}(\boldsymbol{\theta}_0 + \mathbf{a}_+).\end{aligned}\tag{E.2}$$

Since $\mathcal{Q}(\hat{\boldsymbol{\theta}}) \leq \mathcal{Q}(\boldsymbol{\theta}_0 + \mathbf{a}_+)$,

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+)^{\top} \bar{\mathbf{D}}_+ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+) \leq \mathbf{r}(\boldsymbol{\theta}_0 + \mathbf{a}_+) - \mathbf{r}(\hat{\boldsymbol{\theta}}).\tag{E.3}$$

Given $\|\mathbf{a}_+\|_2 = \mathcal{O}_p(1)$ one obtains

$$\begin{aligned}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+)^{\top} \bar{\mathbf{D}}_+ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+) \\ \leq \mathcal{O}_p\left(\frac{1}{nT}\right) \left(\mathcal{O}_p((1 + \sqrt{nT}\|\mathbf{a}_+\|_2)^2) + \mathcal{O}_p((1 + \sqrt{nT}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2)^2) \right).\end{aligned}\tag{E.4}$$

Multiplying (E.4) by nT , using $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+\|_2^2 \mu_{\min}(\bar{\mathbf{D}}_+) \leq (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+)^{\top} \bar{\mathbf{D}}_+ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+)$, Assumption CS*(ii), and the inequality $\sqrt{(x+y)} \leq \sqrt{x} + \sqrt{y}$ yields

$$\sqrt{nT}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{a}_+\|_2 \leq \mathcal{O}_p(1 + \sqrt{nT}\|\mathbf{a}_+\|_2) + \mathcal{O}_p(1 + \sqrt{nT}\|\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}\|_2) = \mathcal{O}_p(1),$$

and, in turn,

$$\sqrt{nT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \bar{\mathbf{D}}_+^{-1} \mathbf{d}_+ + \mathcal{O}_p(1).$$

The final result follows after applying Lemmas E.4, E.5, and E.6, as Assumption AD*(ii).

E.4. Proofs of Lemmas E.1 - E.6

Proof of Lemma E.1. First, write

$$\begin{aligned}\mathbf{Y} &= \alpha_0 \mathbf{Y}_{-1} + \mathbf{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \mathbf{F}_0^{\top} + \boldsymbol{\varepsilon} \\ &= \alpha_0 \mathbf{Y} \mathbf{W} + \alpha_0 \mathbf{y}_0 \boldsymbol{\pi}_T^{\top} + \mathbf{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \mathbf{F}_0^{\top} + \boldsymbol{\varepsilon}.\end{aligned}$$

Thus,

$$\mathbf{Y} = (\alpha_0 \mathbf{y}_0 \boldsymbol{\pi}_T^{\top} + \mathbf{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \mathbf{F}_0^{\top} + \boldsymbol{\varepsilon})(\mathbf{I}_T - \alpha_0 \mathbf{W})^{-1},$$

where $\mathbf{I}_T - \alpha_0 \mathbf{W}$ is invertible under Assumption MD*(i). Using this,

$$\begin{aligned}\mathbf{Y}_{-1} &= \mathbf{Y} \mathbf{W} + \mathbf{y}_0 \boldsymbol{\pi}_T^{\top} \\ &= (\mathbf{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \mathbf{F}_0^{\top} + \boldsymbol{\varepsilon}) \mathbf{G} + \mathbf{y}_0 \boldsymbol{\pi}_T^{\top} (\mathbf{I}_T + \alpha_0 \mathbf{G}) \\ &= (\mathbf{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \mathbf{F}_0^{\top} + \boldsymbol{\varepsilon}) \mathbf{G} + \mathbf{y}_0 \mathbf{g}^{\top} \\ &= \mathbf{H}_1 + \boldsymbol{\Lambda}_0 \mathbf{F}_0^{\top} \mathbf{G} + \boldsymbol{\varepsilon} \mathbf{G}.\end{aligned}$$

□

Proof of Lemma E.1(ii). Since

$$\|\mathbf{g}(\alpha)\|_2 \leq (1 + |\alpha| \times \|\mathbf{G}(\alpha)\|_2) \|\boldsymbol{\pi}_T\|_2,$$

where $|\alpha| \leq c$ under Assumption MD*(i) and $\|\boldsymbol{\pi}_T\|_2 = 1$, it suffices to show that $\|\mathbf{G}(\alpha)\|_2 \leq c$ uniformly over T . Recall $\mathbf{G}(\alpha) = (\mathbf{I}_T - \alpha \mathbf{W})^{-1} \mathbf{W}$, with $\mathbf{W}$ being a $T \times T$ shift matrix of all zeros, except those elements directly above the main diagonal, which equal to 1. Since $\|\mathbf{W}\|_2 = 1$, it remains only to consider $(\mathbf{I}_T - \alpha \mathbf{W})^{-1}$. By Fact 9.13.21 in Bernstein (2009),

$$\|(\mathbf{I}_T - \alpha \mathbf{W})^{-1}\|_1 = \|(\mathbf{I}_T - \alpha \mathbf{W})^{-1}\|_\infty \leq \frac{1}{1 - |\alpha|},$$

using Assumption MD*(i). All results then follow from the inequality $\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty}$. $\square$

Proof of Lemma E.2(i). The lemma is obtained by combining two results. The first is that

$$\|\tilde{\boldsymbol{\varepsilon}}\|_2 = \mathcal{O}_p(T^{\frac{3}{4}}), \quad (\text{E.5})$$

and the second is that

$$\|\boldsymbol{\varepsilon}\|_2 = \mathcal{O}_p\left(\sqrt{\max\{n, T\}}\right). \quad (\text{E.6})$$

Taken together one obtains the result. For (E.5), this follows identically to expression (C.2) in the proof of Lemma B.2(i). For (E.6), under Assumption ED*, the elements of $\boldsymbol{\varepsilon}$ are independent conditional on $\mathcal{E}_{nT}$. Using the result in Latala (2005),

$$\mathbb{E}[\|\boldsymbol{\varepsilon}\|_2 | \mathcal{E}_{nT}] \lesssim \left(\max_{1 \leq i \leq n} \left(\sum_{t=1}^T \mathbb{E}[\varepsilon_{it}^2 | \mathcal{E}_{nT}] \right)^{\frac{1}{2}} + \max_{1 \leq t \leq T} \left(\sum_{i=1}^n \mathbb{E}[\varepsilon_{it}^2 | \mathcal{E}_{nT}] \right)^{\frac{1}{2}} + \left(\sum_{i=1}^n \sum_{t=1}^T \mathbb{E}[\varepsilon_{it}^4 | \mathcal{E}_{nT}] \right)^{\frac{1}{4}} \right),$$

from which it then follows that $\|\boldsymbol{\varepsilon}\|_2 = \mathcal{O}_p(\sqrt{\max\{n, T\}})$. The result is then obtained using $\|\tilde{\boldsymbol{\varepsilon}}\|_2 \leq \|\boldsymbol{\varepsilon}\|_2$ since $\|\mathbf{Q}_{\mathcal{X}}\|_2 = 1$. $\square$

Proof of Lemma E.2(ii). For $k = 1$,

$$\begin{aligned} \|\mathbf{Z}_1\|_F &\leq \|\mathbf{H}_1\|_F + \|\boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G}\|_F + \|\boldsymbol{\varepsilon} \mathbf{G}\|_F \\ &\leq \|\mathbf{H}_1\|_F + \|\boldsymbol{\Lambda}_0\|_F \|\mathbf{F}_0\|_F \|\mathbf{G}\|_2 + \|\boldsymbol{\varepsilon}\|_F \|\mathbf{G}\|_2, \end{aligned} \quad (\text{E.7})$$

using Lemma E.1. Under Assumption AD*(i), similar to the proof of Lemma B.1(i), it can be established that $\|\mathbf{H}_1\|_F = \mathcal{O}_p(\sqrt{nT})$. $\|\boldsymbol{\Lambda}_0\|_F \|\mathbf{F}_0\|_F = \mathcal{O}_p(\sqrt{nT})$ by Lemma B.1(ii), and $\|\mathbf{G}\|_2 = \mathcal{O}(1)$ by Lemma E.1(ii). Finally, it is straightforward to show $\|\boldsymbol{\varepsilon}\|_F = \mathcal{O}_p(\sqrt{nT})$ under Assumption ED*. This gives $\|\mathbf{Z}_1\|_F = \mathcal{O}_p(\sqrt{nT})$, from which follows $\|\tilde{\mathbf{Z}}_1\|_F \leq \|\mathbf{Q}_{\mathcal{X}}\|_2 \|\mathbf{Z}_1\|_F = \|\mathbf{Z}_1\|_F = \mathcal{O}_p(\sqrt{nT})$ and $\|\tilde{\mathbf{Z}}_1\|_2 \leq \|\mathbf{Q}_{\mathcal{X}}\|_2 \|\mathbf{Z}_1\|_2 = \|\mathbf{Z}_1\|_2 \leq \|\mathbf{Z}_1\|_F = \mathcal{O}_p(\sqrt{nT})$. The results for $k = 2, \dots, K+1$ follow by Lemma B.1(i). $\square$

Proof of Lemma E.2(iii). For $k = 1$, using Lemma E.1,

$$\text{tr}(\tilde{\mathbf{Z}}_1^\top \tilde{\boldsymbol{\varepsilon}}) = \text{tr}(\mathbf{H}_1^\top \mathbf{P}_{\mathcal{X}} \boldsymbol{\varepsilon}) + \text{tr}((\boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G})^\top \mathbf{P}_{\mathcal{X}} \boldsymbol{\varepsilon}) + \text{tr}((\boldsymbol{\varepsilon} \mathbf{G})^\top \mathbf{P}_{\mathcal{X}} \boldsymbol{\varepsilon}).$$

Consider each of these terms, in turn. First,

$$\begin{aligned} \mathbb{E} \left[\text{tr}(\mathbf{H}_1^\top \mathbf{P}_{\mathcal{X}} \boldsymbol{\varepsilon})^2 \right] &= \mathbb{E} \left[\left(\sum_{t=1}^T \sum_{i=1}^n (\mathbf{P}_{\mathcal{X}} \mathbf{H}_1)_{it} \varepsilon_{it} \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{t_1=1}^T \sum_{i_1=1}^n \sum_{t_2=1}^T \sum_{i_2=1}^n (\mathbf{P}_{\mathcal{X}} \mathbf{H}_1)_{i_1 t_1} (\mathbf{P}_{\mathcal{X}} \mathbf{H}_1)_{i_2 t_2} \varepsilon_{i_1 t_1} \varepsilon_{i_2 t_2} \right] \\ &\lesssim \mathbb{E} \left[\|\mathbf{P}_{\mathcal{X}} \mathbf{H}_1\|_F^2 \right] \\ &= \mathcal{O}(nT), \end{aligned}$$

using $\|\mathbf{P}_{\mathcal{X}} \mathbf{H}_1\|_F \leq \|\mathbf{H}_1\|_F$, as well as Assumptions ED* and AD*(i). Next,

$$\begin{aligned} \mathbb{E} \left[\text{tr}((\boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G})^\top \mathbf{P}_{\mathcal{X}} \boldsymbol{\varepsilon})^2 \right] &= \mathbb{E} \left[\left(\sum_{t=1}^T \sum_{i=1}^n (\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G})_{it} \varepsilon_{it} \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{t_1=1}^T \sum_{i_1=1}^n \sum_{t_2=1}^T \sum_{i_2=1}^n (\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G})_{i_1 t_1} (\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G})_{i_2 t_2} \varepsilon_{i_1 t_1} \varepsilon_{i_2 t_2} \right] \\ &\lesssim \mathbb{E} \left[\|\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G}\|_F^2 \right] \\ &= \mathcal{O}(nT), \end{aligned}$$

using $\|\mathbf{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0 \mathbf{F}_0^\top \mathbf{G}\|_F \leq \|\boldsymbol{\Lambda}_0 \mathbf{F}_0^\top\|_F \|\mathbf{G}\|_2 \leq \|\boldsymbol{\Lambda}_0\|_F \|\mathbf{F}_0\|_F \|\mathbf{G}\|_2$, Lemma E.1(ii), and Assumptions MD*(ii) and ED*. Finally,

$$\begin{aligned} \mathbb{E}[\text{tr}((\boldsymbol{\varepsilon} \mathbf{G})^\top \mathbf{P}_{\mathcal{X}} \boldsymbol{\varepsilon})^2] &= \mathbb{E}[\text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}}) \text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}}) \text{vec}(\boldsymbol{\varepsilon})] \\ &= \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) (\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}})_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}}))^2 \right. \\ &\quad \left. + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}}) \boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}})) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G}^\top \otimes \mathbf{P}_{\mathcal{X}}) \boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}})) \right] \\ &= \mathcal{O}(nT), \end{aligned}$$

using

$$\sum_{i=1}^n \sum_{t=1}^T (\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}})_{(it)(it)}^2 = \text{tr}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}})^2 = \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}}) \boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}})) = 0$$

and

$$|\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G}^\top \otimes \mathbf{P}_{\mathcal{X}}) \boldsymbol{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}}))| \leq nT \|\boldsymbol{\Sigma}_{\mathcal{E}}\|_2^2 \|\mathbf{G}\|_2^2 = \mathcal{O}(nT),$$

which follows by Lemma E.1(ii) and Assumption ED*. By Markov's inequality, one then obtains $\text{tr}(\tilde{\mathbf{Z}}_1^\top \tilde{\boldsymbol{\varepsilon}}) = \mathcal{O}_p(\sqrt{nT})$. The results for $k = 2, \dots, K+1$ follow similarly. $\square$

Proof of Lemma E.2(iv). The proof is analogous to that of Proposition 1, using Lemmas E.2(i), B.1(ii), E.2(ii), E.2(iii), and Assumption CS*(ii). $\square$

Proof of Lemma E.2(v). The proof is analogous to that of Lemma C.1(i), using Assumptions MD*(ii), ED*, and AD*(i). $\square$

Proof of Lemma E.2(vi). The proof is analogous to that of Lemma C.1(v), using Assumptions MD*(ii), ED*, and AD*(i). $\square$

Proof of Lemma E.2(vii).

$$\begin{aligned}
\mathbb{E} [\|\tilde{\varepsilon} \mathbf{G} \mathbf{F}_0\|_F^2] &= \mathbb{E} [\|\mathbf{P}_{\mathcal{X}} \varepsilon \mathbf{G} \mathbf{F}_0\|_F^2] \\
&= \mathbb{E} \left[\sum_{i_1=1}^n \sum_{r=1}^R \left(\sum_{i_2=1}^n \sum_{t=1}^T (\mathbf{P}_{\mathcal{X}})_{i_1 i_2} \varepsilon_{i_2 t} (\mathbf{G} \mathbf{F}_0)_{tr} \right)^2 \right] \\
&= \mathbb{E} \left[\sum_{i_1=1}^n \sum_{r=1}^R \sum_{i_2=1}^n \sum_{t_1=1}^T \sum_{i_3=1}^n \sum_{t_2=1}^T (\mathbf{P}_{\mathcal{X}})_{i_1 i_2} \varepsilon_{i_2 t_1} (\mathbf{G} \mathbf{F}_0)_{t_1 r} (\mathbf{P}_{\mathcal{X}})_{i_1 i_3} \varepsilon_{i_3 t_2} (\mathbf{G} \mathbf{F}_0)_{t_2 r} \right] \\
&\lesssim T \|\mathbf{G}\|_2^2 \mathbb{E} [\|\mathbf{F}_0\|_F^2] \\
&= \mathcal{O}(T^2),
\end{aligned}$$

which follows using Lemma E.1(ii) and Assumptions MD*(ii) and ED*. $\square$

Proof of Lemma E.3(i). First,

$$\begin{aligned}
\mathbb{E} [\text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\varepsilon}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\varepsilon} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}) | \mathcal{E}_{nT}] &= \mathbb{E} [\text{vec}(\varepsilon)^\top (\mathbf{M}_{\mathbf{F}_0} \mathbf{G}^\top \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\varepsilon) | \mathcal{E}_{nT}] \\
&= \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{M}_{\mathbf{F}_0} \mathbf{G}^\top \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top)) \\
&=: \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E} [\text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\varepsilon}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\varepsilon} \mathbf{P}_{\mathbf{F}_0} \mathbf{G})^2 | \mathcal{E}_{nT}] &= \mathbb{E} [(\text{vec}(\varepsilon)^\top (\mathbf{M}_{\mathbf{F}_0} \mathbf{G}^\top \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\varepsilon))^2 | \mathcal{E}_{nT}] \\
&= \sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) \\
&\quad + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})^2,
\end{aligned}$$

which follows using Assumption ED*. Now,

$$\begin{aligned}
&\mathbb{E} \left[\left(\text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\varepsilon}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\varepsilon} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}) - \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{M}_{\mathbf{F}_0} \mathbf{G}^\top \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top)) \right)^2 \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) \right] \\
&= \mathcal{O}(T),
\end{aligned}$$

since

$$\begin{aligned} \left| \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 \right] \right| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T) \\ |\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})]| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T) \\ |\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top)]| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T), \end{aligned}$$

which follows using Lemma E.1(ii) and Assumption ED*. $\square$

Proof of Lemma E.3(ii). First,

$$\begin{aligned} \mathbb{E} \left[\text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top) | \mathcal{E}_{nT} \right] &= \mathbb{E} \left[\text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{P}_{\mathbf{X}}) \text{vec}(\boldsymbol{\varepsilon}) | \mathcal{E}_{nT} \right] \\ &= \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{P}_{\mathbf{X}})) \\ &=: \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \left[\text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top)^2 | \mathcal{E}_{nT} \right] &= \mathbb{E} \left[(\text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{P}_{\mathbf{X}}) \text{vec}(\boldsymbol{\varepsilon}))^2 | \mathcal{E}_{nT} \right] \\ &= \sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) \\ &\quad + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})^2, \end{aligned}$$

which follows using Assumption ED*. Now,

$$\begin{aligned} &\mathbb{E} \left[\left(\text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{P}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top) - \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \otimes \mathbf{P}_{\mathbf{X}})) \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) \right] \\ &= \mathcal{O}(T), \end{aligned}$$

since

$$\begin{aligned} \left| \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 \right] \right| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T) \\ |\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})]| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T) \\ |\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top)]| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T), \end{aligned}$$

which follows using Lemma E.1(ii) and Assumption ED*. $\square$

Proof of Lemma E.3(iii). Under Assumption ED*,

$$\begin{aligned}
\mathbb{E} \left[\text{tr}(\tilde{\varepsilon} \mathbf{G} \tilde{\varepsilon}^\top \mathbf{P}_{\tilde{\Lambda}_0})^2 \right] &= \mathbb{E} \left[\mathbb{E} \left[(\text{vec}(\varepsilon)^\top (\mathbf{G} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\varepsilon))^2 | \mathcal{E}_{nT} \right] \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) \right] \\
&= \mathcal{O}(T),
\end{aligned}$$

which follows because $\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) = \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) = 0$ and

$$\begin{aligned}
\left| \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 \right] \right| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T) \\
|\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top)]| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T),
\end{aligned}$$

by Lemma E.1(ii) and Assumption ED*. □

Proof of Lemma E.3(iv). First,

$$\begin{aligned}
\mathbb{E} \left[\text{tr}(\tilde{\varepsilon} \mathbf{G} \mathbf{P}_{F_0} \tilde{\varepsilon}^\top \mathbf{P}_{\tilde{\Lambda}_0}) | \mathcal{E}_{nT} \right] &= \mathbb{E} \left[\text{vec}(\varepsilon)^\top (\mathbf{G} \mathbf{P}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\varepsilon) | \mathcal{E}_{nT} \right] \\
&= \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{P}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top)) \\
&=: \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E} \left[\text{tr}(\tilde{\varepsilon} \mathbf{G} \mathbf{P}_{F_0} \tilde{\varepsilon}^\top \mathbf{P}_{\tilde{\Lambda}_0})^2 | \mathcal{E}_{nT} \right] &= \mathbb{E} \left[(\text{vec}(\varepsilon)^\top (\mathbf{G} \mathbf{P}_{F_0} \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\varepsilon))^2 | \mathcal{E}_{nT} \right] \\
&= \sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) \\
&\quad + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})^2,
\end{aligned}$$

which follows using Assumption ED*. Thus

$$\begin{aligned}
&\mathbb{E} \left[\left(\text{tr}(\tilde{\varepsilon} \mathbf{G} \mathbf{P}_{F_0} \tilde{\varepsilon}^\top \mathbf{P}_{\tilde{\Lambda}_0}) - \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) \right)^2 \right] \\
&= \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) \right] \\
&= \mathcal{O}(1),
\end{aligned}$$

since

$$\begin{aligned}
\left| \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 \right] \right| &\lesssim \|\mathbf{G}\|_2^2 = \mathcal{O}(1) \\
|\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})]| &\lesssim \|\mathbf{G}\|_2^2 = \mathcal{O}(1) \\
|\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top)]| &\lesssim \|\mathbf{G}\|_2^2 = \mathcal{O}(1),
\end{aligned}$$

by Lemma E.1(ii) and Assumption ED*. Finally, the result follows from

$$|\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})| \lesssim \|\boldsymbol{\Sigma}_{\mathcal{E}}\|_2 \|\mathbf{G}\|_2 = \mathcal{O}(1),$$

using Lemma E.1(ii) and Assumption ED*. $\square$

Proof of Lemma E.3(v). Let

$$\text{tr}(\tilde{\mathbf{H}}_k \mathbf{M}_{F_0} (\tilde{\mathbf{E}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) = \text{tr}(\boldsymbol{\varepsilon} \mathbf{G} \mathbf{M}_{F_0} \tilde{\mathbf{H}}_k^\top \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) =: \text{tr}(\boldsymbol{\varepsilon} \mathbf{A}).$$

Now,

$$\begin{aligned} \mathbb{E}[\text{tr}(\boldsymbol{\varepsilon} \mathbf{A})^2] &= \mathbb{E} \left[\sum_{i_1=1}^n \sum_{t_1=1}^T \sum_{i_2=1}^n \sum_{t_2=1}^T \varepsilon_{i_1 t_1} \varepsilon_{i_2 t_2} \mathbf{A}_{t_1 i_1} \mathbf{A}_{t_2 i_2} \right] \\ &\lesssim \mathbb{E} [\|\mathbf{A}\|_F^2] \\ &\lesssim \|\mathbf{G}\|_2^2 \mathbb{E} [\|\mathbf{H}_k\|_F^2] \\ &= \mathcal{O}(nT), \end{aligned}$$

using Lemma E.1(ii) and Assumptions ED* and AD*(i). $\square$

Proof of Lemma E.3(vi). First, using Assumption ED*,

$$\begin{aligned} \mathbb{E} \left[\text{tr}(\tilde{\mathbf{E}} \mathbf{G} \mathbf{M}_{F_0} (\tilde{\mathbf{E}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) | \mathcal{E}_{nT} \right] &= \mathbb{E} \left[\text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \mathbf{M}_{F_0} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\boldsymbol{\varepsilon}) | \mathcal{E}_{nT} \right] \\ &= \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{M}_{F_0} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top)) \\ &=: \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \left[\text{tr}(\tilde{\mathbf{E}} \mathbf{G} \mathbf{M}_{F_0} (\tilde{\mathbf{E}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0})^2 | \mathcal{E}_{nT} \right] &= \mathbb{E} \left[(\text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \mathbf{M}_{F_0} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathcal{X}} \mathbf{M}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathcal{X}}^\top) \text{vec}(\boldsymbol{\varepsilon}))^2 | \mathcal{E}_{nT} \right] \\ &= \sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) \\ &\quad + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})^2. \end{aligned}$$

Thus,

$$\begin{aligned} &\mathbb{E} \left[\left(\text{tr}(\tilde{\mathbf{E}} \mathbf{G} \mathbf{M}_{F_0} (\tilde{\mathbf{E}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) - \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top) \right] \\ &= \mathcal{O}(T^2), \end{aligned}$$

using

$$\begin{aligned} \left| \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E},it}^4 - 3\sigma_{\mathcal{E},it}^4) \mathbf{A}_{(it)(it)}^2 \right] \right| &\lesssim T^2 \|\mathbf{G}\|_2^4 \\ |\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A})]| &\lesssim T^2 \|\mathbf{G}\|_2^4 \\ |\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}^\top)]| &\lesssim T^2 \|\mathbf{G}\|_2^4, \end{aligned}$$

by Lemma E.1(ii) and Assumption ED*. Hence,

$$\frac{1}{nT} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{M}_{\mathbf{F}_0} (\tilde{\boldsymbol{\varepsilon}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) = \frac{1}{nT} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) + \mathcal{O}_p\left(\frac{1}{n}\right).$$

Finally, notice that

$$\begin{aligned} \frac{1}{nT} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{A}) &= \frac{1}{nT} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{G}^\top \otimes \mathbf{P}_{\mathbf{X}})) - \frac{1}{nT} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}^\top \otimes \mathbf{P}_{\mathbf{X}})) \\ &\quad - \frac{1}{nT} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathbf{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathbf{X}}^\top)) \\ &\quad + \frac{1}{nT} \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathbf{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathbf{X}}^\top)) \\ &= \frac{1}{nT} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}} (\mathbf{G} \mathbf{G}^\top \otimes \mathbf{I}_{TK})) + \mathcal{O}\left(\frac{1}{n}\right), \end{aligned}$$

using

$$\begin{aligned} |\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathbf{X}} \mathbf{Q}_{\mathbf{X}}^\top))| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T) \\ |\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathbf{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathbf{X}}^\top))| &\lesssim T \|\mathbf{G}\|_2^2 = \mathcal{O}(T) \\ |\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} (\mathbf{G} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}^\top \otimes \mathbf{Q}_{\mathbf{X}} \mathbf{P}_{\tilde{\Lambda}_0} \mathbf{Q}_{\mathbf{X}}^\top))| &\lesssim \|\mathbf{G}\|_2^2 = \mathcal{O}(1), \end{aligned}$$

by Lemma E.1(ii). □

Proof of Lemma E.3(vii). Analogous to the proof of Lemma C.2(i). □

Proof of Lemma E.3(viii). Analogous to the proof of Lemma C.2(ii). □

Proof of Lemma E.4(i). Using Lemma E.1,

$$\begin{aligned}
\mathbf{b}_+^{(1)} &= -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\mathbf{Z}}_1 \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\ \vdots \\ \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\mathbf{Z}}_{K+1} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \end{pmatrix} \\
&= -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\mathbf{H}}_1 \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\ \vdots \\ \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\mathbf{H}}_{K+1} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \end{pmatrix} \\
&\quad - \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\ 0 \\ \vdots \\ 0 \end{pmatrix}.
\end{aligned}$$

For the first term on the right in the above, similar to the proof of Lemma B.3(i),

$$-\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\mathbf{H}}_1 \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \\ \vdots \\ \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\mathbf{H}}_{K+1} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \end{pmatrix} = -\psi^{(1)} + \mathcal{O}_p(n^{-\frac{1}{4}}),$$

using Lemmas B.1(ii), C.1(ii), C.1(iii), E.2(v), and E.3(vii), as well as Assumption AE*. For the second term, consider the nonzero element.

$$\begin{aligned}
&\frac{1}{\sqrt{nT}} \left| \text{tr} \left(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}} \right) \right| \\
&\lesssim \frac{1}{\sqrt{nT}} \frac{1}{nT} \|\tilde{\boldsymbol{\varepsilon}}\|_2 \|\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \right)^{-1} \right\|_2 \|\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_2 \\
&= \mathcal{O}_p \left(n^{-\frac{1}{4}} \right),
\end{aligned}$$

using Lemmas C.1(ii), E.2(i), and E.2(vii), as well as Assumption AE*. Thus,

$$\mathbf{b}_+^{(1)} = -\psi^{(1)} + \mathcal{O}_p(n^{-\frac{1}{4}}).$$

□

Proof of Lemma E.4(ii). Using Lemma E.1,

$$\begin{aligned}
\mathbf{b}_+^{(2)} &= -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{F_0} \tilde{\mathbf{Z}}_1^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \\ \vdots \\ \text{tr} \left(\mathbf{M}_{F_0} \tilde{\mathbf{Z}}_{K+1}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \end{pmatrix} \\
&= -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{F_0} \tilde{\mathbf{H}}_1^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \\ \vdots \\ \text{tr} \left(\mathbf{M}_{F_0} \tilde{\mathbf{H}}_{K+1}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \end{pmatrix} \\
&\quad - \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{F_0} (\tilde{\mathbf{e}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \\ 0 \\ \vdots \\ 0 \end{pmatrix}.
\end{aligned}$$

For the first term on the right in the above, similar to the proof of Lemma B.3(ii),

$$-\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr} \left(\mathbf{M}_{F_0} \tilde{\mathbf{H}}_1^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \\ \vdots \\ \text{tr} \left(\mathbf{M}_{F_0} \tilde{\mathbf{H}}_{K+1}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \end{pmatrix} = \mathcal{O}_p(n^{-\frac{1}{2}}),$$

using Lemmas B.1(ii), C.1(ii), C.1(iv), C.1(vi), and E.2(v), as well as Assumption AE*. For the nonzero element of the second term on the right,

$$\begin{aligned}
&\frac{1}{\sqrt{nT}} \left| \text{tr} \left(\mathbf{M}_{F_0} (\tilde{\mathbf{e}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\mathbf{e}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{e}} \right) \right| \\
&\lesssim \frac{1}{\sqrt{nT}} \frac{1}{nT} \|\tilde{\mathbf{e}}\|_2 \|\mathbf{G}\|_2 \|\tilde{\mathbf{e}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\Lambda}_0^\top \tilde{\Lambda}_0 \right)^{-1} \right\|_2 \|\tilde{\Lambda}_0^\top \tilde{\mathbf{e}}\|_2 \\
&= \mathcal{O}_p \left(n^{-\frac{1}{4}} \right),
\end{aligned}$$

using Lemmas C.1(ii), C.1(vi), E.2(i), and E.1(ii), as well as Assumption AE*. Hence,

$$\mathbf{b}_+^{(2)} = \mathcal{O}_p \left(n^{-\frac{1}{4}} \right).$$

□

Proof of Lemma E.4(iii). Using Lemma E.1,

$$\begin{aligned}
\mathbf{b}_+^{(3)} &= -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{Z}}_1) \\ \vdots \\ \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{Z}}_{K+1}) \end{pmatrix} \\
&= -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{H}}_1) \\ \vdots \\ \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{H}}_{K+1}) \end{pmatrix} \\
&\quad -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\
&\quad -\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{G}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{E.8}
\end{aligned}$$

For the first term on the right in (E.8), similar to the proof of Lemma B.3(iii),

$$-\frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{H}}_1) \\ \vdots \\ \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\mathbf{H}}_{K+1}) \end{pmatrix} = -\psi^{(2)} + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),$$

using Lemmas B.1(ii), C.1(ii), C.1(iii), E.2(v), and E.3(viii), as well as Assumption AE*. For the second term in (E.8), using Lemma E.3(i),

$$\begin{aligned}
-\frac{1}{\sqrt{nT}} \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{P}_{\mathbf{F}_0} \mathbf{G}) &= -\frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\mathbf{P}_{\mathbf{F}_0} \mathbf{G} \mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0})) + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right) \\
&= -\frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\mathbf{P}_{\mathbf{F}_0} \mathbf{G} \mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{I}_{TK})) + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right) \\
&=: -\boldsymbol{\xi}^{(1)} + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

which follows because

$$\frac{1}{\sqrt{nT}} |\text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\mathbf{P}_{\mathbf{F}_0} \mathbf{G} \mathbf{M}_{\mathbf{F}_0} \otimes \mathbf{P}_{\tilde{\Lambda}_0}))| \lesssim \frac{1}{\sqrt{nT}} \|\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}\|_2 \|\mathbf{G}\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{nT}}\right).$$

using Lemma E.1(ii) and Assumption ED*. Finally, for the third term in (E.8),

$$\begin{aligned}
&\frac{1}{\sqrt{nT}} \left| \text{tr}(\mathbf{M}_{\mathbf{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0} \tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0 (\mathbf{F}_0^\top \mathbf{F}_0)^{-1} (\tilde{\Lambda}_0^\top \tilde{\Lambda}_0)^{-1} \tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}} \mathbf{G}) \right|_2 \\
&\lesssim \frac{1}{\sqrt{nT}} \frac{1}{nT} \|\tilde{\boldsymbol{\varepsilon}}\|_2 \|\tilde{\boldsymbol{\varepsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2 \left\| \left(\frac{1}{n} \tilde{\Lambda}_0^\top \tilde{\Lambda}_0 \right)^{-1} \right\|_2 \|\tilde{\Lambda}_0^\top \tilde{\boldsymbol{\varepsilon}}\|_2 \|\mathbf{G}\|_2 \\
&= \mathcal{O}_p\left(n^{-\frac{1}{4}}\right),
\end{aligned}$$

using Lemmas C.1(ii), C.1(vi), E.2(i), and E.1(ii) and Assumption AE*. Thus,

$$\mathbf{b}_+^{(3)} = -\boldsymbol{\psi}^{(2)} - \boldsymbol{\xi}^{(1)} + \mathcal{O}_p\left(n^{-\frac{1}{4}}\right).$$

□

Proof of Lemma E.5(i). Using Lemma E.1,

$$\begin{aligned}\bar{\mathbf{c}}_+ &= \frac{1}{\sqrt{nT}} \tilde{\mathbf{Z}}^\top (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \text{vec}(\tilde{\boldsymbol{\varepsilon}}) \\ &= \frac{1}{\sqrt{nT}} \tilde{\mathbf{H}}^\top (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \text{vec}(\tilde{\boldsymbol{\varepsilon}}) + \frac{1}{\sqrt{nT}} \begin{pmatrix} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{M}_{F_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}.\end{aligned}$$

Consider the second term in the above.

$$\begin{aligned}\frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{M}_{F_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{M}_{\tilde{\Lambda}_0}) &= \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \tilde{\boldsymbol{\varepsilon}}^\top) - \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{P}_{F_0} \tilde{\boldsymbol{\varepsilon}}^\top) \\ &\quad - \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\Lambda}_0}) + \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \mathbf{P}_{F_0} \tilde{\boldsymbol{\varepsilon}}^\top \mathbf{P}_{\tilde{\Lambda}_0}) \\ &= \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \tilde{\boldsymbol{\varepsilon}}^\top) - \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\mathbf{G} \mathbf{P}_{F_0} \otimes \mathbf{I}_{TK})) + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

using Lemmas E.3(ii), E.3(iii), and E.3(iv). Let

$$\begin{aligned}\mathbf{c}_+ &:= \frac{1}{\sqrt{nT}} \tilde{\mathbf{H}}^\top (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \text{vec}(\tilde{\boldsymbol{\varepsilon}}) + \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\varepsilon}} \mathbf{G} \tilde{\boldsymbol{\varepsilon}}^\top) \boldsymbol{\pi}_{K+1} \\ \boldsymbol{\xi}^{(2)} &:= \frac{1}{\sqrt{nT}} \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\mathbf{G} \mathbf{P}_{F_0} \otimes \mathbf{I}_{TK})) \boldsymbol{\pi}_{K+1}.\end{aligned}$$

Then

$$\bar{\mathbf{c}}_+ = \mathbf{c}_+ - \boldsymbol{\xi}^{(2)} + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right).$$

□

Proof of Lemma E.5(ii). Using Lemma E.1,

$$\bar{\mathbf{D}}_+ =: \boldsymbol{\Upsilon}^{(1)} + \mathbf{P}_1 + \mathbf{P}_1^\top + \mathbf{P}_2,$$

where

$$\mathbf{P}_1 := \frac{1}{nT} \begin{pmatrix} \text{tr}(\tilde{\mathbf{H}}_1 \mathbf{M}_{F_0} (\tilde{\boldsymbol{\varepsilon}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) & \cdots & \text{tr}(\tilde{\mathbf{H}}_{K+1} \mathbf{M}_{F_0} (\tilde{\boldsymbol{\varepsilon}} \mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix},$$

and

$$\mathbf{P}_2 := \frac{1}{nT} \begin{pmatrix} \text{tr}(\tilde{\mathbf{E}}\mathbf{G}\mathbf{M}_{F_0}(\tilde{\mathbf{E}}\mathbf{G})^\top \mathbf{M}_{\tilde{\Lambda}_0}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}.$$

Each of these matrices is now examined. First, by Lemma E.3(v), $\mathbf{P}_1 = \mathcal{O}_p((nT)^{-\frac{1}{2}})$. For $\mathbf{P}_2$, by Lemma E.3(vi),

$$\begin{aligned} \mathbf{P}_2 &= \frac{1}{nT} \text{tr}(\tilde{\Sigma}_{\mathcal{E}}(\mathbf{G}\mathbf{G}^\top \otimes \mathbf{I}_{TK}))\boldsymbol{\pi}_{K+1}\boldsymbol{\pi}_{K+1}^\top + \mathcal{O}_p\left(\frac{1}{n}\right) \\ &= \Upsilon^{(2)} + \mathcal{O}_p\left(\frac{1}{n}\right). \end{aligned}$$

Hence,

$$\bar{\mathbf{D}}_+ = \Upsilon^{(1)} + \Upsilon^{(2)} + \mathcal{O}_p\left(\frac{1}{\sqrt{nT}}\right) = \mathbf{D}_+ + \mathcal{O}_p\left(\frac{1}{\sqrt{nT}}\right).$$

□

Proof of Lemma E.6. Let $\mathbf{v}$ denote a nonstochastic $(K+1) \times 1$ vector such that $\|\mathbf{v}\|_2 = 1$. Then

$$\begin{aligned} \mathbf{v}^\top \mathbf{c}_+ &= \frac{1}{\sqrt{nT}} \mathbf{v}^\top \left(\tilde{\mathcal{H}}^\top (\mathbf{M}_{F_0} \otimes \mathbf{M}_{\tilde{\Lambda}_0}) \text{vec}(\tilde{\mathbf{E}}) + \text{tr}(\tilde{\mathbf{E}}\mathbf{G}\tilde{\mathbf{E}}^\top) \boldsymbol{\pi}_{K+1} \right) \\ &= \frac{1}{\sqrt{nT}} \sum_{i_1=1}^n \sum_{t_1=1}^T \left(\mathbf{v}^\top \mathbf{h}_{i_1 t_1}^* + \mathbf{v}^\top \left(\boldsymbol{\pi}_{K+1} \sum_{i_2=1}^n \sum_{t_2=1}^{t_1-1} \mathbf{G}_{t_2 t_1} (\mathbf{P}\mathbf{x})_{i_2 i_1} \varepsilon_{i_2 t_2} \right) \right) \varepsilon_{i_1 t_1} \\ &=: \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T (a_{it} + b_{it}) \varepsilon_{it}. \end{aligned}$$

Now, let $j := n(t-1) + i$ for $1 \leq i \leq n$ and $1 \leq t \leq T$ with $j = 1, \dots, nT =: m$. Define the mapping

$$t(j) := 1 + \lfloor \frac{j-1}{n} \rfloor, \quad i(j) := j - n(t(j)-1),$$

and the shorthand $z_{(j)} := z_{i(j)t(j)}$ for a double-indexed quantity z_{it} . Consider the filtration $\mathcal{F}_j := \sigma(\varepsilon_{11}, \varepsilon_{21}, \dots, \varepsilon_{n1}, \dots, \varepsilon_{1,t(j)-1}, \dots, \varepsilon_{n,t(j)-1}, \varepsilon_{1t(j)}, \dots, \varepsilon_{i(j)t(j)}) \vee \mathcal{E}_{nT}$ for $j \geq 1$ and $\mathcal{F}_0 := \mathcal{E}_{nT}$. Set $\omega_{(j)} := (a_{(j)} + b_{(j)})\varepsilon_{(j)}$. Then

$$\mathbf{v}^\top \mathbf{c}_+ = \frac{1}{\sqrt{m}} \sum_{j=1}^m \omega_{(j)},$$

with $\mathbb{E}[\omega_{(j)} | \mathcal{F}_{j-1}] = 0$, from which it follows that $\{(\omega_{(j)}, \mathcal{F}_j) : 1 \leq j \leq m\}$ forms a martingale difference array. Next, for $\delta > 0$ small, $\mathbb{E}[|\omega_{(j)}|^{2+\delta}] \lesssim (\mathbb{E}[a_{(j)}^4])^{\frac{2+\delta}{4}} + (\mathbb{E}[b_{(j)}^4])^{\frac{2+\delta}{4}}$ using Assumption ED*. Hence, $\mathbb{E}[|\omega_{(j)}|^{2+\delta}] \leq c_\omega$ follows if it can be shown that $\mathbb{E}[a_{it}^4] \leq c_a$ and $\mathbb{E}[b_{it}^4] \leq c_b$. For the first of these,

$$\mathbb{E}[a_{it}^4] \leq \mathbb{E} \left[\left(\sum_{k=1}^{K+1} v_k h_{kit}^* \right)^4 \right] \lesssim \|\mathbf{v}\|_1^4 \leq c, \quad (\text{E.9})$$

using Assumption **AD*(i)** and $\|\mathbf{v}\|_2 = 1$. Next, let $\mathbf{S}_{it}$ be an $n \times T$ matrix of all zeros, except element it , which equals to 1, and define $\mathbf{s}_{it} := \text{vec}(\mathbf{S}_{it})$. Then

$$\begin{aligned}
\mathbb{E}[b_{i_1 t_1}^4] &= \mathbb{E} \left[\left(\mathbf{v}^\top \boldsymbol{\pi}_{K+1} \left(\sum_{i_2=1}^n \sum_{t_2=1}^{t_1-1} G_{t_2 t_1}(\mathbf{P}\mathbf{x})_{i_2 i_1} \varepsilon_{i_2 t_2} \right) \right)^4 \right] \\
&= \mathbb{E} \left[(v_1 \text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \otimes \mathbf{P}\mathbf{x}) \mathbf{s}_{i_1 t_1})^4 \right] \\
&= v_1^4 \mathbb{E} \left[\text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \otimes \mathbf{P}\mathbf{x}) \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top (\mathbf{G}^\top \otimes \mathbf{P}\mathbf{x}) \text{vec}(\boldsymbol{\varepsilon}) \right. \\
&\quad \times \text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \otimes \mathbf{P}\mathbf{x}) \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top (\mathbf{G}^\top \otimes \mathbf{P}\mathbf{x}) \text{vec}(\boldsymbol{\varepsilon}) \left. \right] \\
&=: v_1^4 \mathbb{E} \left[\text{vec}(\boldsymbol{\varepsilon})^\top \mathbf{C} \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top \mathbf{C}^\top \text{vec}(\boldsymbol{\varepsilon}) \text{vec}(\boldsymbol{\varepsilon})^\top \mathbf{C} \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top \mathbf{C}^\top \text{vec}(\boldsymbol{\varepsilon}) \right] \\
&= v_1^4 \mathbb{E} \left[\sum_{i_2=1}^n \sum_{t_2=1}^T (\xi_{\mathcal{E}, i_2 t_2}^4 - 3\sigma_{\mathcal{E}, i_2 t_2}^4) (\mathbf{s}_{i_2 t_2}^\top \mathbf{C} \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top \mathbf{C}^\top \mathbf{s}_{i_2 t_2})^2 \right. \\
&\quad \left. + 2\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top \mathbf{C}^\top \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top \mathbf{C}^\top) + \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top \mathbf{C}^\top)^2 \right] \leq c,
\end{aligned}$$

with $\mathbf{C} := (\mathbf{G} \otimes \mathbf{P}\mathbf{x})$, and which follows because

$$\begin{aligned}
\sum_{i_2=1}^n \sum_{t_2=1}^T (\xi_{\mathcal{E}, i_2 t_2}^4 - 3\sigma_{\mathcal{E}, i_2 t_2}^4) (\mathbf{s}_{i_2 t_2}^\top \mathbf{C} \mathbf{s}_{i_1 t_1} \mathbf{s}_{i_1 t_1}^\top \mathbf{C}^\top \mathbf{s}_{i_2 t_2})^2 &\lesssim \|\mathbf{G}\|_2^4 \\
|\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^\top \mathbf{C}^\top \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^\top \mathbf{C}^\top)| &\lesssim \|\mathbf{G}\|_2^4 \\
\text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^\top \mathbf{C}^\top)^2 &\lesssim \|\mathbf{G}\|_2^4,
\end{aligned}$$

since $\|\mathbf{s}_{it}\|_2 = 1$, and using Lemma **E.1(ii)** and Assumption **ED***. Thus, $\mathbb{E}[b_{it}^4] \leq c_b$ and so $\mathbb{E}[|\omega_{(j)}|^{2+\delta}] \leq c_\omega$ for $\delta > 0$ small. Next,

$$\begin{aligned}
&\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (a_{it} + b_{it})^2 \varepsilon_{it}^2 - \mathbf{v}^\top \mathbf{V}_+ \mathbf{v} \\
&= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T a_{it}^2 (\varepsilon_{it}^2 - \sigma_{\mathcal{E}, it}^2) + \frac{1}{nT} \left(\sum_{i=1}^n \sum_{t=1}^T b_{it}^2 \varepsilon_{it}^2 - v_1^2 \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}} (\mathbf{G} \otimes \mathbf{I}_{TK}) \tilde{\boldsymbol{\Sigma}}_{\mathcal{E}} (\mathbf{G}^\top \otimes \mathbf{I}_{TK})) \right) \\
&\quad + \frac{2}{nT} \sum_{i=1}^n \sum_{t=1}^T a_{it} b_{it} \varepsilon_{it}^2. \tag{E.10}
\end{aligned}$$

Consider each of these terms. For the first term in (E.10),

$$\begin{aligned}
&\mathbb{E} \left[\left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T a_{it}^2 (\varepsilon_{it}^2 - \sigma_{\mathcal{E}, it}^2) \right)^2 \right] \\
&= \mathbb{E} \left[\frac{1}{n^2 T^2} \sum_{i_1=1}^n \sum_{t_1=1}^T \sum_{i_2=1}^n \sum_{t_2=1}^T a_{i_1 t_1}^2 a_{i_2 t_2}^2 (\varepsilon_{i_1 t_1}^2 - \sigma_{\mathcal{E}, i_1 t_1}^2) (\varepsilon_{i_2 t_2}^2 - \sigma_{\mathcal{E}, i_2 t_2}^2) \right] \\
&= \frac{1}{n^2 T^2} \mathbb{E} \left[\sum_{i=1}^n \sum_{t=1}^T (\xi_{\mathcal{E}, it}^4 - \sigma_{\mathcal{E}, it}^4) a_{it}^4 \right] = \mathcal{O}(1),
\end{aligned}$$

using (E.9) and Assumptions ED* and AD*(i). Next, for the second term in (E.10),

$$\begin{aligned}
& \frac{1}{nT} \left(\sum_{i=1}^n \sum_{t=1}^T b_{it}^2 \varepsilon_{it}^2 - v_1^2 \text{tr}(\tilde{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{I}_{TK}) \tilde{\Sigma}_{\mathcal{E}}(\mathbf{G}^{\top} \otimes \mathbf{I}_{TK})) \right) \\
&= \frac{v_1^2}{nT} \left(\sum_{i=1}^n \sum_{t=1}^T \text{vec}(\varepsilon)^{\top} (\mathbf{G} \otimes \mathbf{P}_{\mathcal{X}}) \mathbf{s}_{it} \mathbf{s}_{it}^{\top} (\mathbf{G}^{\top} \otimes \mathbf{P}_{\mathcal{X}}) \text{vec}(\varepsilon) \varepsilon_{it}^2 \right. \\
&\quad \left. - \text{tr}(\tilde{\Sigma}_{\mathcal{E}}(\mathbf{G} \otimes \mathbf{I}_{TK}) \tilde{\Sigma}_{\mathcal{E}}(\mathbf{G}^{\top} \otimes \mathbf{I}_{TK})) \right) \\
&= \frac{v_1^2}{nT} \left(\sum_{i=1}^n \sum_{t=1}^T \text{vec}(\varepsilon)^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \text{vec}(\varepsilon) \varepsilon_{it}^2 - \text{tr}(\Sigma_{\mathcal{E}} \mathbf{C} \Sigma_{\mathcal{E}} \mathbf{C}^{\top}) \right) \\
&= \frac{v_1^2}{nT} \left(\sum_{i=1}^n \sum_{t=1}^T (\mathbf{e}_{-it} + \varepsilon_{it} \mathbf{s}_{it})^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} (\mathbf{e}_{-it} + \varepsilon_{it} \mathbf{s}_{it}) \varepsilon_{it}^2 - \text{tr}(\Sigma_{\mathcal{E}} \mathbf{C} \Sigma_{\mathcal{E}} \mathbf{C}^{\top}) \right),
\end{aligned}$$

where $\mathbf{e}_{-it} := \text{vec}(\varepsilon) - \varepsilon_{it} \mathbf{s}_{it}$. Now, notice,

$$\begin{aligned}
& \sum_{i=1}^n \sum_{t=1}^T (\mathbf{e}_{-it} + \varepsilon_{it} \mathbf{s}_{it})^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} (\mathbf{e}_{-it} + \varepsilon_{it} \mathbf{s}_{it}) \varepsilon_{it}^2 \\
&= \sum_{i=1}^n \sum_{t=1}^T \mathbf{e}_{-it}^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{e}_{-it} \varepsilon_{it}^2 + 2 \sum_{i=1}^n \sum_{t=1}^T \mathbf{e}_{-it}^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{s}_{it} \varepsilon_{it}^3 \\
&\quad + \sum_{i=1}^n \sum_{t=1}^T \mathbf{s}_{it}^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{s}_{it} \varepsilon_{it}^4 \\
&= \sum_{i=1}^n \sum_{t=1}^T \mathbf{e}_{-it}^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{e}_{-it} \varepsilon_{it}^2,
\end{aligned}$$

which follows since the second and third term in the second display above are exactly zero as $\mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{s}_{it}$ corresponds to a diagonal element of $\mathbf{C}$. Continuing,

$$\begin{aligned}
& \mathbb{E} \left[\left(\sum_{i=1}^n \sum_{t=1}^T \mathbf{e}_{-it}^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{e}_{-it} \varepsilon_{it}^2 - \text{tr}(\Sigma_{\mathcal{E}} \mathbf{C} \Sigma_{\mathcal{E}} \mathbf{C}^{\top}) \right)^2 \right] \\
&= \mathbb{E} \left[\left(\sum_{i=1}^n \sum_{t=1}^T \mathbf{e}_{-it}^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{e}_{-it} \varepsilon_{it}^2 \right)^2 - \text{tr}(\Sigma_{\mathcal{E}} \mathbf{C} \Sigma_{\mathcal{E}} \mathbf{C}^{\top})^2 \right]. \tag{E.11}
\end{aligned}$$

Next,

$$\begin{aligned}
& \mathbb{E} \left[\left(\sum_{i=1}^n \sum_{t=1}^T \mathbf{e}_{-it}^{\top} \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^{\top} \mathbf{C}^{\top} \mathbf{e}_{-it} \varepsilon_{it}^2 \right)^2 \middle| \mathcal{E}_{nT} \right] \\
&=: \mathbb{E} \left[\left(\sum_{j=1}^m \mathbf{e}_{(-j)}^{\top} \mathbf{C} \mathbf{s}_{(j)} \mathbf{s}_{(j)}^{\top} \mathbf{C}^{\top} \mathbf{e}_{(-j)} \varepsilon_{(j)}^2 \right)^2 \middle| \mathcal{E}_{nT} \right] \\
&= \mathbb{E} \left[\sum_{j=1}^m \mathbf{e}_{(-j)}^{\top} \mathbf{C} \mathbf{s}_{(j)} \mathbf{s}_{(j)}^{\top} \mathbf{C}^{\top} \mathbf{e}_{(-j)} \mathbf{e}_{(-j)}^{\top} \mathbf{C} \mathbf{s}_{(j)} \mathbf{s}_{(j)}^{\top} \mathbf{C}^{\top} \mathbf{e}_{(-j)} \varepsilon_{(j)}^4 \middle| \mathcal{E}_{nT} \right] \\
&\quad + \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \mathbf{e}_{(-j_1)}^{\top} \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{s}_{(j_1)}^{\top} \mathbf{C}^{\top} \mathbf{e}_{(-j_1)} \varepsilon_{(j_1)}^2 \mathbf{e}_{(-j_2)}^{\top} \mathbf{C} \mathbf{s}_{(j_2)} \mathbf{s}_{(j_2)}^{\top} \mathbf{C}^{\top} \mathbf{e}_{(-j_2)} \varepsilon_{(j_2)}^2 \middle| \mathcal{E}_{nT} \right] \\
&=: p_1 + p_2.
\end{aligned}$$

Consider each of these terms.

$$p_1 = \sum_{j_1=1}^m \sum_{j_2=1}^m \xi_{\mathcal{E},(j_1)}^4 (\xi_{\mathcal{E},(j_2)}^4 - 3\sigma_{\mathcal{E},(j_2)}^4) \mathbf{C}_{j_2 j_1}^4 + 3 \sum_{j_1=1}^m \sum_{j_2=1}^m \sum_{j_3=1}^m \xi_{\mathcal{E},(j_1)}^4 \sigma_{\mathcal{E},(j_2)}^2 \sigma_{\mathcal{E},(j_3)}^2 \mathbf{C}_{j_2 j_1}^2 \mathbf{C}_{j_3 j_1}^2$$

$$=: p_{1.1} + p_{1.2}.$$

Notice

$$|p_{1.1}| \lesssim \|\mathbf{C}\|_{\max}^2 \|\mathbf{C}\|_F^2 \lesssim T^2 \|\mathbf{G}\|_2^4 = \mathcal{O}(T^2)$$

$$|p_{1.2}| \lesssim \|\mathbf{C}\|_{\infty,2}^2 \|\mathbf{C}\|_F^2 \lesssim T^2 \|\mathbf{G}\|_2^4 = \mathcal{O}(T^2),$$

which follows because

$$\|\mathbf{C}\|_{\max} \leq \|\mathbf{G}\|_{\max} \|\mathbf{P}_{\mathcal{X}}\|_{\max} \leq \|\mathbf{G}\|_2 \leq c,$$

and

$$\|\mathbf{C}\|_{\infty,2}^2 \leq \max_{\mathbf{v}: \|\mathbf{v}\|_2=1} \|\mathbf{C}\mathbf{v}\|_2^2 = \|\mathbf{C}\|_2^2 \leq \|\mathbf{G}\|_2^2 \leq c,$$

using $\|\mathbf{P}_{\mathcal{X}}\|_{\max} \leq \|\mathbf{P}_{\mathcal{X}}\|_2 = 1$ and $\|\mathbf{G}\|_{\max} \leq \|\mathbf{G}\|_2 \leq c$ by Lemma E.1(ii). Thus, $p_1 = \mathcal{O}(T^2)$. Next,

$$p_2 = \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m (\mathbf{e}_{(-j_1-j_2)} + \varepsilon_{(j_2)} \mathbf{s}_{(j_2)})^\top \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{s}_{(j_1)}^\top \mathbf{C}^\top (\mathbf{e}_{(-j_1-j_2)} + \varepsilon_{(j_2)} \mathbf{s}_{(j_2)}) \varepsilon_{(j_1)}^2 \right.$$

$$\left. \times (\mathbf{e}_{(-j_1-j_2)} + \varepsilon_{(j_1)} \mathbf{s}_{(j_1)})^\top \mathbf{C} \mathbf{s}_{(j_2)} \mathbf{s}_{(j_2)}^\top \mathbf{C}^\top (\mathbf{e}_{(-j_1-j_2)} + \varepsilon_{(j_1)} \mathbf{s}_{(j_1)}) \varepsilon_{(j_2)}^2 \right] \mathcal{E}_{nT}$$

$$=: p_{2.1} + \dots + p_{2.9}. \quad (\text{E.12})$$

For the first term in (E.12),

$$p_{2.1} = \sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3 \neq j_1, j_2}^m \sigma_{\mathcal{E},j_1}^2 \sigma_{\mathcal{E},j_2}^2 (\xi_{\mathcal{E},j_3}^4 - 3\sigma_{\mathcal{E},j_3}^4) C_{j_3 j_1}^2 C_{j_3 j_2}^2$$

$$+ 2 \sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3 \neq j_1, j_2}^m \sum_{j_4 \neq j_1, j_2}^m \sigma_{\mathcal{E},j_1}^2 \sigma_{\mathcal{E},j_2}^2 \sigma_{\mathcal{E},j_3}^2 \sigma_{\mathcal{E},j_4}^2 C_{j_3 j_1} C_{j_3 j_2} C_{j_4 j_1} C_{j_4 j_2}$$

$$+ \sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3 \neq j_1, j_2}^m \sum_{j_4 \neq j_1, j_2}^m \sigma_{\mathcal{E},j_1}^2 \sigma_{\mathcal{E},j_2}^2 \sigma_{\mathcal{E},j_3}^2 \sigma_{\mathcal{E},j_4}^2 C_{j_3 j_1}^2 C_{j_4 j_2}^2$$

$$=: p_{2.1.1} + p_{2.1.2} + p_{2.1.3}.$$

Now, first, $|p_{2.1.1}| \lesssim \|\mathbf{C}\|_{2,\infty} \|\mathbf{C}\|_F^2 \leq T^2 \|\mathbf{G}\|_2^4 = \mathcal{O}(T^2)$. Second,

$$p_{2.1.2} = 2 \sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3 \neq j_1, j_2}^m \sum_{j_4 \neq j_1, j_2}^m \sigma_{\mathcal{E},j_1}^2 \sigma_{\mathcal{E},j_2}^2 \sigma_{\mathcal{E},j_3}^2 \sigma_{\mathcal{E},j_4}^2 C_{j_3 j_1} C_{j_3 j_2} C_{j_4 j_1} C_{j_4 j_2}$$

$$= 2 \text{tr}(\mathbf{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{\Sigma}_{\mathcal{E}} \mathbf{C}^\top \mathbf{\Sigma}_{\mathcal{E}} \mathbf{C} \mathbf{\Sigma}_{\mathcal{E}} \mathbf{C}^\top) - 2 \sum_{j_1=1}^m \sum_{j_2=1}^m \sum_{j_3=1}^m \sigma_{\mathcal{E},j_1}^4 \sigma_{\mathcal{E},j_2}^2 \sigma_{\mathcal{E},j_3}^2 C_{j_2 j_1}^2 C_{j_3 j_1}^2$$

$$=: p_{2.1.2.1} + p_{2.1.2.2},$$

which follows using $C_{jj} = 0$. Notice that

$$\begin{aligned} |p_{2.1.2.1}| &\lesssim \text{rank}(\mathbf{C}) \|\boldsymbol{\Sigma}_{\mathcal{E}}\|_2^4 \|\mathbf{C}\|_2^4 \lesssim T^2 \|\mathbf{G}\|_2^4 = \mathcal{O}(T^2), \\ |p_{2.1.2.2}| &\lesssim \|\mathbf{C}\|_{2,\infty} \|\mathbf{C}\|_F^2 \leq T^2 \|\mathbf{G}\|_2^4 = \mathcal{O}(T^2), \end{aligned}$$

and thereby $p_{2.1.2} = \mathcal{O}(T^2)$. Third,

$$\begin{aligned} p_{2.1.3} &= \sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3 \neq j_1, j_2}^m \sum_{j_4 \neq j_1, j_2}^m \sigma_{\mathcal{E},j_1}^2 \sigma_{\mathcal{E},j_2}^2 \sigma_{\mathcal{E},j_3}^2 \sigma_{\mathcal{E},j_4}^2 C_{j_3 j_1}^2 C_{j_4 j_2}^2 \\ &= \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C}^\top)^2 - \sum_{j_1=1}^m \sum_{j_2=1}^m \sum_{j_3=1}^m \sigma_{\mathcal{E},j_1}^4 \sigma_{\mathcal{E},j_2}^2 \sigma_{\mathcal{E},j_3}^2 C_{j_2 j_1}^2 C_{j_3 j_1}^2 \\ &\quad - \sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3=1}^m \sigma_{\mathcal{E},j_1}^2 \sigma_{\mathcal{E},j_2}^4 \sigma_{\mathcal{E},j_3}^2 C_{j_2 j_1}^2 C_{j_3 j_2}^2 - \sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3 \neq j_1, j_2}^m \sigma_{\mathcal{E},j_1}^4 \sigma_{\mathcal{E},j_2}^2 \sigma_{\mathcal{E},j_3}^2 C_{j_3 j_1}^2 C_{j_1 j_2}^2 \\ &=: \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C}^\top)^2 + p_{2.1.3.1} + p_{2.1.3.2} + p_{2.1.3.3}. \end{aligned}$$

Notice

$$|p_{2.1.3.1}| \lesssim \|\mathbf{C}\|_{2,\infty} \|\mathbf{C}\|_F^2 \leq T^2 \|\mathbf{G}\|_2^4 = \mathcal{O}(T^2).$$

Similarly for $p_{2.1.3.2}$ and $p_{2.1.3.3}$, whereby $p_{2.1.3} = \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C}^\top)^2 + \mathcal{O}(T^2)$ and, in turn, $p_{2.1} = \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C}^\top)^2 + \mathcal{O}(T^2)$. Returning to (E.12), in similar fashion one can analyse terms $p_{2.1} + \dots + p_{2.9}$ and establish these to be $\mathcal{O}(T^2)$. It therefore follows that

$$\mathbb{E} \left[\left(\sum_{i=1}^n \sum_{t=1}^T \mathbf{e}_{-it}^\top \mathbf{C} \mathbf{s}_{it} \mathbf{s}_{it}^\top \mathbf{C}^\top \mathbf{e}_{-it} \varepsilon_{it}^2 - \text{tr}(\boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C} \boldsymbol{\Sigma}_{\mathcal{E}} \mathbf{C}^\top) \right)^2 \right] = \mathcal{O}(T^2),$$

whereby the second term in (E.10) is $\mathcal{O}_p(1)$. For the third term in (E.10),

$$\begin{aligned} \frac{2}{nT} \sum_{i=1}^n \sum_{t=1}^T a_{it} b_{it} \varepsilon_{it}^2 &= \frac{2v_1}{nT} \sum_{i=1}^n \sum_{t=1}^T a_{it} \text{vec}(\boldsymbol{\varepsilon})^\top (\mathbf{G} \otimes \mathbf{P}_{\mathbf{x}}) \mathbf{s}_{it} \varepsilon_{it}^2 \\ &= \frac{2v_1}{m} \sum_{j=1}^m a_{(j)} (\mathbf{e}_{(-j)} + \mathbf{s}_{(j)} \varepsilon_{(j)})^\top \mathbf{C} \mathbf{s}_{(j)} \varepsilon_{(j)}^2 \\ &= \frac{2v_1}{m} \sum_{j=1}^m a_{(j)} \mathbf{e}_{(-j)}^\top \mathbf{C} \mathbf{s}_{(j)} \varepsilon_{(j)}^2, \end{aligned}$$

because $\mathbf{C}$ has a zero diagonal. Now,

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{2v_1}{m} \sum_{j=1}^m a_{(j)} \mathbf{e}_{(-j)}^\top \mathbf{C} \mathbf{s}_{(j)} \varepsilon_{(j)}^2 \right)^2 \right] \\
&= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2=1}^m a_{(j_1)} a_{(j_2)} \mathbf{e}_{(-j_1)}^\top \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{e}_{(-j_2)}^\top \mathbf{C} \mathbf{s}_{(j_2)} \varepsilon_{(j_1)}^2 \varepsilon_{(j_2)}^2 \right] \\
&= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j=1}^m a_{(j)}^2 \mathbf{e}_{(-j)}^\top \mathbf{C} \mathbf{s}_{(j)} \mathbf{s}_{(j)}^\top \mathbf{C}^\top \mathbf{e}_{(-j)} \varepsilon_{(j)}^4 \right] \\
&\quad + \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} \mathbf{e}_{(-j_1)}^\top \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{e}_{(-j_2)}^\top \mathbf{C} \mathbf{s}_{(j_2)} \varepsilon_{(j_1)}^2 \varepsilon_{(j_2)}^2 \right] \\
&=: q_1 + q_2.
\end{aligned}$$

Consider each of these terms in turn.

$$\begin{aligned}
q_1 &= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \xi_{\mathcal{E},(j_1)}^4 \sigma_{\mathcal{E},(j_2)}^2 a_{(j_1)}^2 \mathbf{C}_{j_2 j_1}^2 \right] \\
&\lesssim \frac{1}{m^2} \mathbb{E} \left[\left(\max_{1 \leq j_1 \leq m} a_{(j_1)}^2 \right) \|\mathbf{C}\|_F^2 \right] \\
&\lesssim \frac{T^2}{m^2} \mathbb{E} \left[\max_{1 \leq j_1 \leq m} a_{(j_1)}^2 \right] \\
&\lesssim \frac{T^2}{m^2} \left(\mathbb{E} \left[\max_{1 \leq j_1 \leq m} a_{(j_1)}^4 \right] \right)^{\frac{1}{2}} \\
&\lesssim \frac{T^2}{m^2} \left(\mathbb{E} \left[\sum_{j_1=1}^m a_{(j_1)}^4 \right] \right)^{\frac{1}{2}} = o(1),
\end{aligned}$$

using Assumptions **ED*** and **AD*(i)**. Next,

$$\begin{aligned}
q_2 &= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} (\mathbf{e}_{(-j_1-j_2)} + \mathbf{s}_{(j_2)} \varepsilon_{(j_2)})^\top \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{s}_{(j_2)}^\top \mathbf{C}^\top (\mathbf{e}_{(-j_1-j_2)} + \mathbf{s}_{(j_1)} \varepsilon_{(j_1)}) \varepsilon_{(j_1)}^2 \varepsilon_{(j_2)}^2 \right] \\
&= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} \mathbf{e}_{(-j_1-j_2)}^\top \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{s}_{(j_2)}^\top \mathbf{C}^\top \mathbf{e}_{(-j_1-j_2)} \varepsilon_{(j_1)}^2 \varepsilon_{(j_2)}^2 \right] \\
&\quad + \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} \mathbf{C}_{j_1 j_2} \mathbf{e}_{(-j_1-j_2)}^\top \mathbf{C} \mathbf{s}_{(j_1)} \varepsilon_{(j_1)}^3 \varepsilon_{(j_2)}^2 \right] \\
&\quad + \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} \mathbf{C}_{j_2 j_1} \mathbf{s}_{(j_2)}^\top \mathbf{C}^\top \mathbf{e}_{(-j_1-j_2)} \varepsilon_{(j_1)}^2 \varepsilon_{(j_2)}^3 \right] \\
&\quad + \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} \mathbf{C}_{j_2 j_1} \mathbf{C}_{j_1 j_2} \varepsilon_{(j_1)}^3 \varepsilon_{(j_2)}^3 \right] \\
&= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} \mathbf{e}_{(-j_1-j_2)}^\top \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{s}_{(j_2)}^\top \mathbf{C}^\top \mathbf{e}_{(-j_1-j_2)} \varepsilon_{(j_1)}^2 \varepsilon_{(j_2)}^2 \right],
\end{aligned}$$

since $\mathbb{E}[\mathbf{e}_{(-j_1-j_2)}\varepsilon_{(j_1)}^2\varepsilon_{(j_2)}^3|\mathcal{E}_{nT}] = \mathbf{0}$ and $\mathbf{C}_{j_2j_1}\mathbf{C}_{j_1j_2} = 0$, and where $\mathbf{e}_{(-j_1-j_2)} := \text{vec}(\boldsymbol{\varepsilon}) - \varepsilon_{(j_1)}\mathbf{s}_{(j_1)} - \varepsilon_{(j_2)}\mathbf{s}_{(j_2)}$. Thus,

$$\begin{aligned}
q_2 &= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m a_{(j_1)} a_{(j_2)} \mathbf{e}_{(-j_1-j_2)}^\top \mathbf{C} \mathbf{s}_{(j_1)} \mathbf{s}_{(j_2)}^\top \mathbf{C}^\top \mathbf{e}_{(-j_1-j_2)} \varepsilon_{(j_1)}^2 \varepsilon_{(j_2)}^2 \right] \\
&= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2 \neq j_1}^m \sum_{j_3 \neq j_1, j_2}^m \sigma_{\varepsilon, (j_1)}^2 \sigma_{\varepsilon, (j_2)}^2 \sigma_{\varepsilon, (j_3)}^2 a_{(j_1)} a_{(j_2)} \mathbf{C}_{j_3j_1} \mathbf{C}_{j_3j_2} \right] \\
&= \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2=1}^m \sum_{j_3=1}^m \sigma_{\varepsilon, (j_1)}^2 \sigma_{\varepsilon, (j_2)}^2 \sigma_{\varepsilon, (j_3)}^2 a_{(j_1)} a_{(j_2)} \mathbf{C}_{j_3j_1} \mathbf{C}_{j_3j_2} \right] \\
&\quad - \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2=1}^m \sigma_{\varepsilon, (j_1)}^4 \sigma_{\varepsilon, (j_2)}^2 a_{(j_1)}^2 \mathbf{C}_{j_2j_1}^2 \right] \\
&=: q_{2.1} + q_{2.2}.
\end{aligned}$$

For the first term,

$$\begin{aligned}
q_{2.1} &= \frac{4v_1^2}{m^2} \mathbb{E}[\mathbf{a}^\top \boldsymbol{\Sigma}_\varepsilon \mathbf{C}^\top \boldsymbol{\Sigma}_\varepsilon \mathbf{C} \boldsymbol{\Sigma}_\varepsilon \mathbf{a}] \\
&\leq \frac{4v_1^2}{m^2} \mathbb{E} \left[\|\mathbf{a}\|_2^2 \|\boldsymbol{\Sigma}_\varepsilon \mathbf{C}^\top \boldsymbol{\Sigma}_\varepsilon \mathbf{C} \boldsymbol{\Sigma}_\varepsilon\|_2 \right] \\
&\lesssim \frac{4v_1^2}{m^2} \mathbb{E} [\|\mathbf{a}\|_2^2] = \mathcal{O}(1).
\end{aligned}$$

using Assumptions [ED*](#) and [AD*\(i\)](#). Next,

$$|q_{2.2}| = \frac{4v_1^2}{m^2} \mathbb{E} \left[\sum_{j_1=1}^m \sum_{j_2=1}^m \sigma_{\varepsilon, (j_1)}^4 \sigma_{\varepsilon, (j_2)}^2 a_{(j_1)}^2 \mathbf{C}_{j_2j_1}^2 \right] = \mathcal{O}(1),$$

similar to q_1 . Thus, $q_2 = \mathcal{O}(1)$, whereby the third term in [\(E.10\)](#) is $\mathcal{O}_p(1)$, and overall

$$\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (a_{it} + b_{it})^2 \varepsilon_{it}^2 - \mathbf{v}^\top \mathbf{V}_+ \mathbf{v} = \mathcal{O}_p(1),$$

Hence, with $\mathbf{v}^\top \mathbf{V}_+ \mathbf{v} > 0$ w.p.a.1 by Assumption [AD*\(ii\)](#), the result follows by the Cramér-Wold device and Corollary 5.26 in [White \(1984\)](#). $\square$

Appendix F. Higher Order Terms

This appendix studies the order four and order five terms detailed in the proof of Proposition [2](#).

F.1. Order Four

Define

$$\mathcal{E}^{(4)} := \frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}). \quad (\text{F.1})$$

The summation in (F.1) can be decomposed into 16 terms by separating the index 0 and the remaining indices which are greater than 0. This gives

$$\begin{aligned}
\mathcal{E}^{(4)} &= \frac{1}{nT} e_0 e_0 e_0 e_0 \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
&+ \frac{1}{nT} \sum_{k_1=1}^K e_{k_1} e_0 e_0 e_0 \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
&+ \frac{1}{nT} \sum_{k_1=1}^K e_0 e_{k_1} e_0 e_0 \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\
&+ \frac{1}{nT} \sum_{k_1=1}^K e_0 e_0 e_{k_1} e_0 \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0) \\
&+ \frac{1}{nT} \sum_{k_1=1}^K e_0 e_0 e_0 e_{k_1} \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}) \\
&+ 10 \text{ similar terms} \\
&+ \frac{1}{nT} \sum_{k_1=1}^K \sum_{k_2=1}^K \sum_{k_3=1}^K \sum_{k_4=1}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} \mathcal{Q}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}) \\
&=: \mathcal{E}_1^{(4)} + \dots + \mathcal{E}_{16}^{(4)}. \tag{F.2}
\end{aligned}$$

For terms $\mathcal{E}_6^{(4)}, \dots, \mathcal{E}_{16}^{(4)}$, the bound obtained in equation S.58 in Moon and Weidner (2015) can be applied to demonstrate that

$$\mathcal{E}_6^{(4)} + \dots + \mathcal{E}_{16}^{(4)} = \mathcal{O}_p(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^4) + \mathcal{O}_p\left(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^3 \frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}}\right) + \mathcal{O}_p\left(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2 \left(\frac{\|\tilde{\boldsymbol{\varepsilon}}\|_2}{\sqrt{nT}}\right)^2\right).$$

What remains are terms $\mathcal{E}_2^{(4)}, \mathcal{E}_3^{(4)}, \mathcal{E}_4^{(4)}$ and $\mathcal{E}_5^{(4)}$. To evaluate these it is necessary to obtain an explicit expression for

$$\tilde{\mathcal{Q}}^{(4)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}) = \sum_{p=1}^4 (-1)^{p+1} \sum_{\substack{v_1 + \dots + v_p = 4 \\ m_1 + \dots + m_{p+1} = p-1 \\ 1 \leq v_j \leq 2 \\ m_j \geq 0}} \text{tr}(\mathbf{S}^{(m_1)} \boldsymbol{\tau}_{k_1 \dots}^{(v_1)} \mathbf{S}^{(m_2)} \dots \boldsymbol{\tau}_{\dots k_4}^{(v_p)} \mathbf{S}^{(m_{p+1})}).$$

For $p = 1$ there are zero terms, for $p = 2$ there are 3 terms, for $p = 3$ there are 30 terms, and for $p = 4$ there are 35 terms. However, using $\mathbf{S}^{(0)} \mathbf{S}^{(1)} = \mathbf{S}^{(1)} \mathbf{S}^{(0)} = \mathbf{S}^{(0)} \mathbf{S}^{(2)} = \mathbf{S}^{(2)} \mathbf{S}^{(0)} = \mathbf{S}^{(0)} \boldsymbol{\tau}_{k_1}^{(1)} \mathbf{S}^{(0)} = \mathbf{0}$ and the invariance of the matrix trace to cyclical permutations the number of terms can be reduced to 1, 6 and 4 for p equal to 2, 3 and 4, respectively. Going further, several of these terms are equivalent up to

a reordering of the indices k_1, k_2, k_3, k_4 . Therefore the remaining terms are (up to a constant)

$$\begin{aligned}
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathbf{S}^{(1)} \mathcal{T}_{k_3, k_4}^{(2)} \mathbf{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathbf{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathbf{S}^{(2)} \mathcal{T}_{k_4}^{(1)} \mathbf{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathbf{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathbf{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_2, k_3}^{(2)} \mathbf{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathbf{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_3, k_4}^{(2)} \mathbf{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathbf{S}^{(2)} \mathcal{T}_{k_2}^{(1)} \mathbf{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathbf{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathbf{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathbf{S}^{(0)}),
\end{aligned}$$

which are denoted $p_1(k_1, k_2, k_3, k_4), \dots, p_7(k_1, k_2, k_3, k_4)$. Consider the first term.

$$|p_1(k_1, k_2, k_3, k_4)| \lesssim \frac{1}{nT^2} \|\tilde{\mathbf{X}}_{k_1}^\top \tilde{\mathbf{X}}_{k_2} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2^2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\mathbf{F}_0^\top \tilde{\mathbf{X}}_{k_3}^\top \tilde{\mathbf{X}}_{k_4}\|_2.$$

This gives,

$$\begin{aligned}
|p_1(k_1, 0, 0, 0)| &\leq \frac{1}{nT^2} \|\tilde{\mathbf{X}}_{k_1}\|_2 \|\tilde{\mathbf{\epsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2^2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\mathbf{F}_0^\top \tilde{\mathbf{\epsilon}}^\top\|_2 \|\tilde{\mathbf{\epsilon}}\|_2 = \mathcal{O}_p \left(T \frac{\|\tilde{\mathbf{\epsilon}}\|_2}{\sqrt{nT}} \right), \\
|p_1(0, k_2, 0, 0)| &\leq \frac{1}{nT^2} \|\tilde{\mathbf{\epsilon}}^\top \tilde{\mathbf{X}}_{k_2}\|_2 \|\mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2^2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\mathbf{F}_0^\top \tilde{\mathbf{\epsilon}}^\top\|_2 \|\tilde{\mathbf{\epsilon}}\|_2 = \mathcal{O}_p \left(T \frac{\|\tilde{\mathbf{\epsilon}}\|_2}{\sqrt{nT}} \right), \\
|p_1(0, 0, k_3, 0)| &\leq \frac{1}{nT^2} \|\tilde{\mathbf{\epsilon}}\|_2 \|\tilde{\mathbf{\epsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2^2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\mathbf{F}_0\|_2 \|\tilde{\mathbf{X}}_{k_3}^\top \tilde{\mathbf{\epsilon}}\|_2 = \mathcal{O}_p \left(T \frac{\|\tilde{\mathbf{\epsilon}}\|_2}{\sqrt{nT}} \right), \\
|p_1(0, 0, 0, k_4)| &\leq \frac{1}{nT^2} \|\tilde{\mathbf{\epsilon}}\|_2 \|\tilde{\mathbf{\epsilon}} \mathbf{F}_0\|_2 \left\| \left(\frac{1}{T} \mathbf{F}_0^\top \mathbf{F}_0 \right)^{-1} \right\|_2^2 \left\| \left(\frac{1}{n} \tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \right)^{-1} \right\|_2 \|\mathbf{F}_0^\top \tilde{\mathbf{\epsilon}}^\top\|_2 \|\tilde{\mathbf{X}}_{k_4}\|_2 = \mathcal{O}_p \left(T \frac{\|\tilde{\mathbf{\epsilon}}\|_2}{\sqrt{nT}} \right),
\end{aligned}$$

using Lemmas [B.1\(i\)](#), [B.1\(ii\)](#), [C.1\(vi\)](#), [C.1\(vii\)](#), and [C.1\(viii\)](#), as well as Assumption [AE](#). Applying Lemma [B.2\(i\)](#) then yields the result $|p_1(k_1, k_2, k_3, k_4)| = \mathcal{O}_p(\sqrt{nT})$. Repeatedly using Lemmas [B.2\(i\)](#) and [C.1](#), as well as Assumption [AE](#), the remaining terms $p_2(k_1, k_2, k_3, k_4), \dots, p_7(k_1, k_2, k_3, k_4)$ can be shown to share the same order. With this result, it follows that

$$\mathcal{E}_2^{(4)}, \mathcal{E}_3^{(4)}, \mathcal{E}_4^{(4)}, \mathcal{E}_5^{(4)} = \mathcal{O}_p \left(\frac{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2}{\sqrt{nT}} \right),$$

and thereby

$$\begin{aligned}
\mathcal{E}^{(4)} &= \frac{1}{nT} e_0 e_0 e_0 e_0 \mathcal{Q}^{(4)}(\tilde{\mathbf{\Lambda}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) + \mathcal{O}_p(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^4) + \mathcal{O}_p \left(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^3 \frac{\|\tilde{\mathbf{\epsilon}}\|_2}{\sqrt{nT}} \right) \\
&+ \mathcal{O}_p \left(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2 \frac{\|\tilde{\mathbf{\epsilon}}\|_2^2}{nT} \right) + \mathcal{O}_p \left(\frac{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2}{\sqrt{nT}} \right).
\end{aligned}$$

F.2. Order Five

As in the previous part write

$$\mathcal{E}^{(5)} := \frac{1}{nT} \sum_{k_1=0}^K \sum_{k_2=0}^K \sum_{k_3=0}^K \sum_{k_4=0}^K \sum_{k_5=0}^K e_{k_1} e_{k_2} e_{k_3} e_{k_4} e_{k_5} Q^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}, \tilde{\mathbf{X}}_{k_5}).$$

This summation can be decomposed into 32 terms by, again, separating out the index 0 and the remaining indices which are greater than 0. This gives

$$\begin{aligned} \mathcal{E}^{(5)} &= \frac{1}{nT} e_0 e_0 e_0 e_0 e_0 Q^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\ &+ \frac{1}{nT} \sum_{k_1=1}^K e_{k_1} e_0 e_0 e_0 e_0 Q^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\ &+ \frac{1}{nT} \sum_{k_1=1}^K e_0 e_{k_1} e_0 e_0 e_0 Q^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\ &+ \frac{1}{nT} \sum_{k_1=1}^K e_0 e_0 e_{k_1} e_0 e_0 Q^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) \\ &+ \frac{1}{nT} \sum_{k_1=1}^K e_0 e_0 e_0 e_{k_1} e_0 Q^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_0) \\ &+ \frac{1}{nT} \sum_{k_1=1}^K e_0 e_0 e_0 e_0 e_{k_1} Q^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_{k_1}) \\ &+ 26 \text{ similar terms} \\ &=: \mathcal{E}_1^{(5)} + \mathcal{E}_2^{(5)} + \dots + \mathcal{E}_{32}^{(5)}. \end{aligned} \tag{F.3}$$

For terms $\mathcal{E}_7^{(5)}, \dots, \mathcal{E}_{32}^{(5)}$, the bound obtained in equation S.58 in [Moon and Weidner \(2015\)](#) can be applied to demonstrate that

$$\begin{aligned} \mathcal{E}_7^{(5)} + \dots + \mathcal{E}_{32}^{(5)} &= \mathcal{O}_p(\|\beta - \beta_0\|_2^5) + \mathcal{O}_p\left(\|\beta - \beta_0\|_2^4 \frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}}\right) \\ &+ \mathcal{O}_p\left(\|\beta - \beta_0\|_2^3 \left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}}\right)^2\right) + \mathcal{O}_p\left(\|\beta - \beta_0\|_2^2 \left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}}\right)^3\right). \end{aligned}$$

What remains are terms $\mathcal{E}_2^{(5)}, \mathcal{E}_3^{(5)}, \mathcal{E}_4^{(5)}, \mathcal{E}_5^{(5)}$ and $\mathcal{E}_6^{(5)}$. As previously it is necessary to make explicit

$$\tilde{Q}^{(5)}(\tilde{\mathbf{A}}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_{k_1}, \tilde{\mathbf{X}}_{k_2}, \tilde{\mathbf{X}}_{k_3}, \tilde{\mathbf{X}}_{k_4}, \tilde{\mathbf{X}}_{k_5}) = \sum_{p=1}^5 (-1)^{p+1} \sum_{\substack{v_1 + \dots + v_p = 5 \\ m_1 + \dots + m_{p+1} = p-1 \\ 1 \leq v_j \leq 2 \\ m_j \geq 0}} \text{tr}(\mathbf{S}^{(m_1)} \mathcal{T}_{k_1 \dots}^{(v_1)} \mathbf{S}^{(m_2)} \dots \mathcal{T}_{\dots k_5}^{(v_p)} \mathbf{S}^{(m_{p+1})}).$$

For $p = 1$ and $p = 2$ there are zero terms, for $p = 3$ there are 30 terms, for $p = 4$ there are 140 terms, and for $p = 5$ there are 126 terms. As before number of terms can be reduced to 9, 22 and 10 for p equal to 3, 4 and 5, respectively. Moreover, since many of these terms are equivalent up to a reordering

of their indices those that remain are (up to a constant)

$$\begin{aligned}
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(0)} \mathcal{T}_{k_3, k_4}^{(2)} \mathcal{S}^{(2)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_4, k_5}^{(2)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(1)} \mathcal{T}_{k_3, k_4}^{(2)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4, k_5}^{(2)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4, k_5}^{(2)} \mathcal{S}^{(1)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(2)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(0)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(0)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(0)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(2)}) \\
& \text{tr}(\mathcal{T}_{k_1, k_2}^{(2)} \mathcal{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2, k_3}^{(2)} \mathcal{S}^{(0)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2}^{(2)} \mathcal{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_3, k_4}^{(2)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4, k_5}^{(2)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathcal{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(0)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathcal{S}^{(0)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(2)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}) \\
& \text{tr}(\mathcal{T}_{k_1}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_2}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_3}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_4}^{(1)} \mathcal{S}^{(1)} \mathcal{T}_{k_5}^{(1)} \mathcal{S}^{(0)}).
\end{aligned}$$

Applying Lemmas B.2(i) and C.1, as well as Assumption AE, it can be established that

$$\mathcal{E}_2^{(5)}, \mathcal{E}_3^{(5)}, \mathcal{E}_4^{(5)}, \mathcal{E}_5^{(5)}, \mathcal{E}_6^{(5)} = \mathcal{O}_p \left(\frac{\|\beta - \beta_0\|_2}{\sqrt{nT}} \right),$$

and thereby

$$\begin{aligned}
\mathcal{E}^{(5)} &= \frac{1}{nT} e_0 e_0 e_0 e_0 e_0 Q^{(5)}(\tilde{\Lambda}_0, \mathbf{F}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_0) + \mathcal{O}_p(\|\beta - \beta_0\|_2^5) + \mathcal{O}_p \left(\|\beta - \beta_0\|_2^4 \frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right) \\
&+ \mathcal{O}_p \left(\|\beta - \beta_0\|_2^3 \left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right)^2 \right) + \mathcal{O}_p \left(\|\beta - \beta_0\|_2^2 \left(\frac{\|\tilde{\epsilon}\|_2}{\sqrt{nT}} \right)^3 \right) + \mathcal{O}_p \left(\frac{\|\beta - \beta_0\|_2}{\sqrt{nT}} \right).
\end{aligned}$$

Appendix G. Additional Monte Carlo Material

Assume, for simplicity, that $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\mathbf{X}_1, \dots, \mathbf{X}_K$ are fixed. Then the largest singular value of $\tilde{\varepsilon} = \mathbf{Q}_\mathbf{X}^\top \varepsilon$ is the same as that of a $\min\{n, TK\} \times T$ matrix with independent $\mathcal{N}(0, \sigma_\varepsilon^2)$ entries.⁴⁴ Therefore,

$$\frac{\|\tilde{\varepsilon}\|_2}{\sqrt{T}} \xrightarrow{p} \left(1 + \sqrt{\min\left\{\frac{1}{\gamma}, K\right\}}\right) \sigma_\varepsilon,$$

as $n, T \rightarrow \infty$ with $T/n \rightarrow \gamma \in [0, \infty)$, with the convention that $1/\gamma = \infty$ when $\gamma = 0$ (see, e.g., Theorem 5.11 in [Bai et al. \(2010\)](#)). Hence,

$$\varrho_n = \left(1 + \sqrt{\frac{\min\{n, TK\}}{T}}\right) \frac{\sigma_\varepsilon}{\sqrt{n}}$$

supplies a penalty of the order required by Proposition 10. Based on this heuristic, the following choice of ϱ_n is considered as a means to tailor the choice of penalty parameter to the data. First, the model is estimated with $R_e = 5 > 2 = R_0$ to obtain an initial estimate $\check{\beta}$. Using this, $\check{\mathbf{A}}$ and $\check{\mathbf{F}}$ are constructed as $\check{\mathbf{A}} := \sqrt{n}(\mathbf{u}_1, \dots, \mathbf{u}_{R_e})$ and $\check{\mathbf{F}} := (s_1 \mathbf{v}_1, \dots, s_{R_e} \mathbf{v}_{R_e})/\sqrt{n}$, respectively, where $\mathbf{Y} - \mathbf{X} \cdot \check{\beta} =: \sum_{t=1}^T s_t \mathbf{u}_t \mathbf{v}_t^\top$ is a singular value decomposition with singular values $s_T \leq \dots \leq s_1$. Finally, with

$$\check{\sigma}_\varepsilon^2 = \frac{1}{nT} \left\| \mathbf{Y} - \mathbf{X} \cdot \check{\beta} - \check{\mathbf{A}} \check{\mathbf{F}}^\top \right\|_F^2,$$

the penalty parameter is set as

$$\check{\varrho}_n = \left(1 + \sqrt{\frac{\min\{n, TK\}}{T}}\right) \frac{\check{\sigma}_\varepsilon}{\sqrt{n}}.$$

Tables G.1 and G.2 below display counterparts to Tables 3 and 6 with $\varrho_n = \check{\varrho}_n$.

Table G.1: Percentage of Estimated R equal to R_0 , $\varrho_n = \check{\varrho}_n$

$n \setminus T$	9	12	15
100	77.44	84.08	88.55
250	90.23	95.09	97.70
500	96.12	98.72	99.44

⁴⁴Recall $\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^\top$, where the matrices of singular vectors $\mathbf{U}$ and $\mathbf{V}$ are $n \times \text{rank}(\mathbf{X})$ and $TK \times \text{rank}(\mathbf{X})$, respectively, and the diagonal matrix of singular values $\mathbf{S}$ is $\text{rank}(\mathbf{X}) \times \text{rank}(\mathbf{X})$. Since $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_{\text{rank}(\mathbf{X})}$, $\|\tilde{\varepsilon}\|_2 = \|\mathbf{V} \mathbf{U}^\top \varepsilon\|_2 = \|\mathbf{U}^\top \varepsilon\|_2$, and $\mathbf{U}^\top \varepsilon$ is a $\min\{n, TK\} \times T$ matrix with independent $\mathcal{N}(0, \sigma_\varepsilon^2)$ entries.

Table G.2: Percentage of Estimated R equal to R_0 , $\varrho_n = \check{\varrho}_n$

$n \setminus T$	10	25	50
100	87.03	98.88	99.94
250	95.38	99.99	100.00
500	98.20	100.00	100.00

Comparing Tables G.1 and G.2 to Tables 3 and 6, the data-driven choice of ϱ_n results in a small improvement in test performance.

References

- Bai, Z., Silverstein, J. W., et al., 2010. Spectral analysis of large dimensional random matrices. Vol. 20. Springer.
- Bernstein, D. S., 2009. Matrix mathematics: theory, facts, and formulas. Princeton University Press, Princeton, New Jersey, USA.
- Hall, P., Heyde, C. C., 1980. Martingale Limit Theory and Its Application. Probability and Mathematical Statistics. Academic Press, New York.
- Latala, R., 2005. Some estimates of norms of random matrices. Proceedings of the American Mathematical Society 133 (5), 1273–1282.
- Lin, Z., 2010. Probability inequalities. Springer.
- Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B., Tropp, J. A., 2014. Matrix concentration inequalities via the method of exchangeable pairs. The Annals of Probability 42 (3), 906–945.
- Moon, H. R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83 (4), 1543–1579.
- Moon, H. R., Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. Econometric Theory 33 (1), 158–195.
- White, H., 1984. Asymptotic Theory for Econometricians, 1st Edition. Economic Theory, Econometrics, and Mathematical Economics. Academic Press, San Diego.