Online Supplementary Material for "Post-Selection Inference in Three-Dimensional Panel Data"*

Harold D. Chiang† Joel Rodrigue‡ Yuya Sasaki§

Abstract

This supplementary material contains additional materials. Appendix B presents auxiliary lemmas for the theoretical results, consisting of oracle inequalities (Appendix B.1), a concentration inequality (Appendix B.2), regularized events (Appendix B.3), rates of nuisance parameters (Appendix B.4), sufficiency for Assumption 1 (Appendix B.5–B.7), and empirical pre-sparsity (Appendix B.8). Appendix D presents additional simulation results based on alternative sample sizes (Appendix D.1) and alternative simulation designs (Appendix D.2).

*First arXiv version: March 30, 2019. We benefited from very useful comments by Peter C. B. Phillips (editor), Liangjun Su (co-editor), three anonymous referees, Antonio Galvao, Hiro Kasahara, Kengo Kato, Carlos Lamarche, Whitney Newey, participants in 2019 Cemmap/WISE Workshop on Advances in Econometrics and 2019 University of Tokyo Workshop on Advances in Econometrics. All remaining errors are ours. Code files are available upon request from the authors. Chiang is supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation.

†Harold D. Chiang: hdchiang@wisc.edu. Department of Economics, University of Wisconsin-Madison, William H. Sewell Social Science Building 1180 Observatory Drive Madison, WI 53706-1393, USA

‡Joel Rodrigue: joel.b.rodrigue@vanderbilt.edu. Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA

§Yuya Sasaki: yuya.sasaki@vanderbilt.edu. Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA

Throughout this appendix, denote X^{-l} for the matrix X without its l-th column. Recall that

$$S = \begin{bmatrix} \sqrt{NM} I_{k_0} & 0 & 0 \\ 0 & \sqrt{M} I_{N_0} & 0 \\ 0 & 0 & \sqrt{N} I_{M_0} \end{bmatrix}.$$

Denote $Q = S/\sqrt{NM}$ and $a = p \vee (NM)$. Also, for a matrix A, denote $\|A\|_\infty = \max_{i,j} |A_{i,j}|$.

B Auxiliary Lemmas

B.1 Oracle Inequalities

Assumption B.1 (Oracle Inequalities)**.** *For each (N, M) and for some choice of μ that depends on (N, M), we have $2\|\widehat{\Upsilon}_1^{-1}\varepsilon' X\|_\infty \leq \mu/c$, $2\|\widehat{\Upsilon}_2^{-1}\varepsilon' D_1\|_\infty \leq \mu/\sqrt{N}c$ and $2\|\widehat{\Upsilon}_3^{-1}\varepsilon' D_2\|_\infty \leq \mu/\sqrt{M}c$ with probability $1 - o(1)$ for some $c > 1$.*

Assumption B.2 (Weights for Penalty)**.** *There exists the ideal penalty loading matrix $\widehat{\Upsilon}_l^0$ with all elements bounded and bounded away from zero uniformly over (N, M), sequences u, ℓ with $0 < \ell \leq 1 \leq u$, $\ell \overset{p}{\to} 1$, and $u \overset{p}{\to} u' > 1$ for some constant u' such that*

$$\ell \widehat{\Upsilon}_l^0 \leq \widehat{\Upsilon}_l \leq u \widehat{\Upsilon}_l^0$$

with probability $1 - o(1)$ for $l = 1, 2, 3$.

Remark B.1. *There are many possible situations where one may want to impose weights to penalize different parameters differently. These situations include (1) the case where one incorporates extra information from economic theory; (2) a penalty choice based on the theory of moderate deviation inequality for self-normalized sums as in Belloni, Chen, Chernozhukov and Hansen (2012); (3) the case where one conducts an iterating lasso algorithm such as the conservative lasso as in Caner and Kock (2018); and (4) the common practice of normalizing the standard errors of all covariates to one.* △

For a $p \times p$ matrix A, define the restricted eigenvalue as

$$\kappa_C^2(A, s_1, s_2, s_3) = \min_{\substack{R_1 \subset [k], |R_1| \le s_1 \\ R_2 \subset [N_0], |R_2| \le s_2 \\ R_3 \subset [M_0], |R_3| \le s_3 \\ R = R_1 \cup R_2 \cup R_3}} \min_{\substack{\delta \in \mathbb{R}^p \setminus \{0\} \\ \|\delta_J^c\|_1 \le C \|\delta_J\|_1}} (s_1 + s_2 + s_3) \frac{\delta' A \delta}{\|\delta_J\|_1^2}. \tag{B.1}$$

Assumption B.3 (Restricted Eigenvalues)**.** *For any $C > 0$, there exists $\underline{\kappa}_C > 0$ which depends only on C such that $\kappa_C^2 := \kappa_C^2(\bar{\Psi}, s_1, s_2, s_3) \ge \underline{\kappa}_C$ for all (N, M).*

Remark B.2. *As highlighted in Belloni et al. (2012), Assumption 4 implies that Assumption B.3 holds with probability at least $1 - o(1)$ by the argument in Bickel, Ritov and Tsybakov (2009).* $\triangle$

The following lemma presents oracle inequalities for the three-dimensional panel lasso. Its proof is closely related to Lemma 6 of Belloni et al. (2012). The main difference is that it accounts for the presence of fixed effects with different effective sample sizes.

Lemma B.1 (Oracle Inequalities)**.** *If Assumptions 2, B.1, B.2, and B.3 are satisfied, then*

$$\begin{aligned}
&\|Z(\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}})\| = \sqrt{(\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}})' Q^{-1} \bar{\Psi} Q^{-1} (\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}})} \lesssim \frac{\mu \sqrt{s}}{\sqrt{NM} \kappa_{c_0}} + c_s, \\
&\sqrt{(\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}})' \frac{Z'Z}{NM} (\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}})} \lesssim \frac{\mu \sqrt{s}}{NM \kappa_{c_0}} + \frac{c_s}{\sqrt{NM}}, \\
&\|\widehat{\Upsilon}_1^0 (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_1 \lesssim \frac{\mu s}{NM \kappa_{2c_0} \kappa_{c_0}} + \frac{\sqrt{s} c_s}{\sqrt{NM} \kappa_{2c_0}} + \frac{c_s^2}{\mu}, \\
&\|\widehat{\Upsilon}_2^0 (\widehat{\boldsymbol{\alpha}} - \overline{\boldsymbol{\alpha}})\|_1 \lesssim \frac{\mu s}{\sqrt{NM} \kappa_{2c_0} \kappa_{c_0}} + \frac{\sqrt{s} c_s}{\sqrt{M} \kappa_{2c_0}} + \frac{N^{1/2} c_s^2}{\mu}, \qquad \textit{and} \\
&\|\widehat{\Upsilon}_3^0 (\widehat{\boldsymbol{\gamma}} - \overline{\boldsymbol{\gamma}})\|_1 \lesssim \frac{\mu s}{N \sqrt{M} \kappa_{2c_0} \kappa_{c_0}} + \frac{\sqrt{s} c_s}{\sqrt{N} \kappa_{2c_0}} + \frac{M^{1/2} c_s^2}{\mu}.
\end{aligned}$$

Furthermore, these bounds are valid uniformly over the ℓ_0-ball $\{\overline{\eta} \in \mathbb{R}^p : \|\overline{\eta}\|_0 \le s\}$.

Proof. From the definition of $\widehat{\boldsymbol{\eta}}$, we have

$$\|y - Z\widehat{\boldsymbol{\eta}}\|^2 + \mu P(\widehat{\boldsymbol{\eta}}) \le \|y - Z\overline{\boldsymbol{\eta}}\|^2 + \mu P(\overline{\boldsymbol{\eta}}).$$

Rewriting this inequality yields

$$\begin{aligned}
&\|Z(\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}) + (R + \varepsilon)\|^2 + \mu P(\widehat{\boldsymbol{\eta}}) \le \|(R + \varepsilon)\|^2 + \mu P(\overline{\boldsymbol{\eta}}) \qquad \text{and} \\
&\|Z(\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}) + (R + \varepsilon)\|^2 \le \|(R + \varepsilon)\|^2 + \mu(P(\overline{\boldsymbol{\eta}}) - P(\widehat{\boldsymbol{\eta}})).
\end{aligned}$$

Using the reverse triangle inequality and the dual norm inequality,

$$
\begin{aligned}
\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|^2 \leq & 2|\varepsilon' Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})| + 2|R' Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})| + \mu\Big\{P((\overline{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}})_J) - P((\overline{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}})_{J^c})\Big\} \\
\leq & 2\|(\widehat{\Upsilon}_1)^{-1}\varepsilon' X\|_\infty \|\widehat{\Upsilon}_1(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|_1 + 2\|(\widehat{\Upsilon}_2)^{-1}\varepsilon' D_1\|_\infty \|\widehat{\Upsilon}_2(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})\|_1 \\
& + 2\|(\widehat{\Upsilon}_3)^{-1}\varepsilon' D_2\|_\infty \|\widehat{\Upsilon}_3\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}}\|_1 \\
& + 2\|R\|\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| + \mu\Big\{P((\overline{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}})_J) - P((\overline{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}})_{J^c})\Big\} \\
\leq & \frac{\mu}{c}\Big(\|\widehat{\Upsilon}_1(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})\|_1\Big) \\
& + 2c_s\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| + \mu\Big\{P((\overline{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}})_J) - P((\overline{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}})_{J^c})\Big\},
\end{aligned}
$$

where the third inequality follows from Assumptions 2 and B.1. By the definition of P, we have

$$
\begin{aligned}
\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|^2 \leq & \mu\Big(u+\frac{1}{c}\Big)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big) \\
& - \mu\Big(\ell-\frac{1}{c}\Big)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big) \\
& + 2c_s\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| \qquad \text{(B.2)}
\end{aligned}
$$

under Assumption B.2.

We now branch into two cases. First, suppose that $\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| < 2c_s$. In this case, since $\mu\sqrt{s}/(\sqrt{NM}\kappa_{c_0}) > 0$, we have $\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| \lesssim \mu\sqrt{s}/(\sqrt{NM}\kappa_{c_0}) + c_s$, as required. Second, suppose that $\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| \geq 2c_s$. In this case,

$$
\begin{aligned}
\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|^2 \leq & \mu\Big(u+\frac{1}{c}\Big)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big) \\
& - \mu\Big(\ell-\frac{1}{c}\Big)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big) \\
& + \|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|^2,
\end{aligned}
$$

and thus

$$
\begin{aligned}
& \Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big) \\
\leq & c_0\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big), \qquad \text{(B.3)}
\end{aligned}
$$

where $c_0 = (uc+1)/(\ell c-1)$. Assumption B.3 implies that, for any δ which is in the choice set of the minimum of restricted eigenvalue definition, we have

$$\kappa_{c_0}^2 = \min_{\substack{R_1\subset[k],|R_1|\le s_1\\ R_2\subset[N_0],|R_2|\le s_2\\ R_3\subset[M_0],|R_3|\le s_3\\ R=R_1\cup R_2\cup R_3}} \min_{\substack{\delta\in\mathbb{R}^p\setminus\{0\}\\ \|\delta_J^c\|_1\le C\|\delta_J\|_1}} (s_1+s_2+s_3)\frac{\delta'\bar{\Psi}\delta}{\|\delta_J\|_1^2} \ge \underline{\kappa} > 0.$$

Since $\delta'\Psi\delta = \delta' S^{-1}Z'ZS^{-1}\delta = b'Z'Zb$ for $b = S^{-1}\delta$, we can rewrite the condition in terms of b and obtain

$$\kappa_{c_0}^2 = \min_{\substack{R_1\subset[k],|R_1|\le s_1\\ R_2\subset[N_0],|R_2|\le s_2\\ R_3\subset[M_0],|R_3|\le s_3\\ R=R_1\cup R_2\cup R_3}} \min_{\substack{b\in\mathbb{R}^p\setminus\{0\}\\ \|b^1_{R_1^c}\|_1+\frac{1}{\sqrt{N}}\|b^2_{R_2^c}\|_1+\frac{1}{\sqrt{M}}\|b^3_{R_3^c}\|_1\\ \le c_0\|b_{R_2}\|_1+\frac{1}{\sqrt{N}}\|b^2_{R_2}\|_1+\frac{1}{\sqrt{M}}\|b^3_{R_3}\|_1}} (s_1+s_2+s_3)\frac{\|Zb\|^2}{NM\|(NM)^{-1}Sb_J\|_1^2}.$$

Note that (B.3) implies that we can let $b = \widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}$. Thus,

$$\left\|\begin{matrix}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\\ \frac{1}{\sqrt{N}}(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\\ \frac{1}{\sqrt{M}}(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\end{matrix}\right\|_1^2 \le \frac{(s_1+s_2+s_3)}{\kappa_{c_0}^2 NM}\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|^2.$$

Taking the square root on both sides yields

$$\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1 \le \frac{\sqrt{s_1+s_2+s_3}}{\kappa_{c_0}\sqrt{NM}}\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|. \quad \text{(B.4)}$$

Finally, substituting this equation into (B.2) and dropping the negative terms on the right-hand side yields

$$\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| \le \mu\Big(u+\frac{1}{c}\Big)\frac{\sqrt{s_1+s_2+s_3}}{\kappa_{c_0}\sqrt{NM}} + 2c_s.$$

This shows the first equation in the statement of the lemma. The second result of the lemma follows by dividing the first by $\sqrt{NM}$.

We next obtain the L^1-norm bounds. We branch into two cases. First, suppose that

$$\begin{aligned}&\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big)\\ \le& 2c_0\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big).\end{aligned}$$

By definition of κ_{2c_0}, we have

$$\begin{aligned}
&\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})\|_1\Big)\\
\le&(1+2c_0)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big)\\
\le&(1+2c_0)\frac{\sqrt{s_1+s_2+s_3}}{\kappa_{2c_0}\sqrt{NM}}\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|
\end{aligned}$$

by applying similar lines of arguments to those of the first part of the proof using $2c_0$ in place of c_0. Second, suppose that

$$\begin{aligned}
&\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big)\\
>&2c_0\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big). \qquad \text{(B.5)}
\end{aligned}$$

In this case, equation (B.2) implies

$$\begin{aligned}
\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|^2\le&\mu\Big(u+\frac{1}{c}\Big)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big)\\
&-\mu\Big(\ell-\frac{1}{c}\Big)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big)\\
&+2c_s\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|\le 2c_s\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|,
\end{aligned}$$

where the last inequality is due to the definition of $c_0=(uc+1)/(\ell c-1)$. Equation (B.2) further implies that

$$\begin{aligned}
&\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_1^c}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big)\\
\le&c_0\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big)\\
&+\frac{c}{\ell c-1}\frac{1}{\mu}\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|(2c_s-\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|)\\
\le&c_0\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3}\|_1\Big)\\
&+\frac{c}{\ell c-1}\frac{1}{\mu}c_s^2\\
\le&\frac{c_0}{2}\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1+\frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1+\frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big)\\
&+\frac{c}{\ell c-1}\frac{1}{\mu}c_s^2,
\end{aligned}$$

where the first inequality follows from (B.2), the second inequality follows from $\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|(2c_s - \|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\|) \leq \max_{x\geq 0} x(2c_s - x) \leq c_s^2$, and the third inequality follows from (B.5). Therefore,

$$
\begin{aligned}
&\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})\|_1\Big)\\
\leq&\Big(1+\frac{1}{2c_0}\Big)\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{J_1^c}\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})_{J_2^c}\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})_{J_3^c}\|_1\Big)\\
\leq&\Big(1+\frac{1}{2c_0}\Big)\frac{2c}{\ell c-1}\frac{1}{\mu}c_s^2,
\end{aligned}
$$

where the first inequality is due to (B.5) and the second inequality is due to the previous equation. Combining the two cases together, we obtain

$$
\begin{aligned}
&\Big(\|\widehat{\Upsilon}_1^0(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|_1 + \frac{1}{\sqrt{N}}\|\widehat{\Upsilon}_2^0(\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\alpha}})\|_1 + \frac{1}{\sqrt{M}}\|\widehat{\Upsilon}_3^0(\widehat{\boldsymbol{\gamma}}-\overline{\boldsymbol{\gamma}})\|_1\Big)\\
\leq&(1+2c_0)\frac{\sqrt{s_1+s_2+s_3}}{\kappa_{2c_0}\sqrt{NM}}\|Z(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\| + \Big(1+\frac{1}{2c_0}\Big)\frac{2c}{\ell c-1}\frac{1}{\mu}c_s^2,
\end{aligned}
$$

which, after some re-arrangements, implies results 3-5 in Lemma B.1.

Finally, the uniformity follows from the fact that all the arguments above depend on $\overline{\eta}$ only through s. □

B.2 Concentration Inequality

The following lemma follows from Chernozhukov, Chetverikov and Kato (2014) and Chernozhukov, Chetverikov and Kato (2015).

Lemma B.2 (A Concentration Inequality). *Let $(X_i)_{i\in[n]}$ be p-dimensional independent random vectors, $B=\sqrt{E[\max_{i\in[n]}\|X_i\|_\infty^2]}$, and $\sigma^2=\max_{j\in[p]}\frac{1}{n}\sum_{i=1}^n E|X_{ij}|^2$. With probability at least $1-C(\log n)^{-1}$,*

$$
\max_{j\in[p]}\Big|\frac{1}{n}\sum_{i=1}^n(X_{ij}-E[X_{ij}])\Big| \lesssim \sqrt{\frac{\sigma^2\log(p\vee n)}{n}} + \frac{B\log(p\vee n)}{n}.
$$

Proof. The claim follows from applying Theorem 5.1 of Chernozhukov, Chetverikov and Kato (2014) to Lemma 8 of Chernozhukov, Chetverikov and Kato (2015) with $t=\log n$, $\alpha=1$, and $q=2$. □

B.3 Regularized Events

Lemma B.3 (Regularized Events)**.** *Fix constants $c > 1$ and $C > 0$, and let $\widehat{\Upsilon} = I$. If Assumption 3 is satisfied, then we have $2\|\varepsilon' X\|_\infty \leq \mu/c$, $2\|\varepsilon' D_1\|_\infty \leq \mu/c\sqrt{N}$ and $2\|\varepsilon' D_2\|_\infty \leq \mu/c\sqrt{M}$ with probability at least $1 - C(\log(N \wedge M))^{-1}$ where $\mu = C\sqrt{NM \log a}$. Similarly, if Assumption 3 is satisfied, then we have $\|X^{-l}\zeta^l\|_\infty \leq \mu^l_{node}/2c$, $\|D_1\zeta^l\|_\infty \leq \mu^l_{node}/2c\sqrt{N}$, and $\|D_2\zeta^l\|_\infty \leq \mu_{node,l}/2c\sqrt{M}$ uniformly over $l \in [p]$ with probability at least $1 - C(\log(N \wedge M))^{-1}$ where $\mu^l_{node} = C\sqrt{NM \log a}$.*

Proof. Applying Lemma B.2, we have

$$\begin{aligned}\frac{\|X'\varepsilon\|_\infty}{NM} &= \max_{l\in[k_0]} \Big| \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{t=1}^{T} (X_{ijt,l}\varepsilon_{ijt} - E[X_{ijt,l}\varepsilon_{ijt}]) \Big| \\ &\lesssim \sqrt{\frac{\sigma^2 \log(p \vee (NM))}{NM}} + \frac{B \log(p \vee (NM))}{NM}\end{aligned}$$

with probability $1 - C(\log NM)^{-1}$, where $\sigma^2 = \max_{l\in[k_0]} \max_{t\in[T]} \frac{1}{NM} E[X^2_{ijt,l}\varepsilon^2_{ijt}] \leq O(K^4)$. Note that we have

$$\begin{aligned}B^2 =& E[\max_{i\in[N],j\in[M],t\in[T]} \|X_{ijt}\varepsilon_{ijt}\|^2_\infty] \\ \leq& (E[\max_{i\in[N],j\in[M],t\in[T]} \|X_{ijt}\|^q_\infty |\varepsilon_{ijt}|^q])^{2/q} \\ \lesssim& (NM)^{2/q} (E[\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{t=1}^{T} \|X_{ijt}\|^q_\infty |\varepsilon_{ijt}|^q])^{2/q} \\ \lesssim& (NM)^{2/q} \Big[\Big(E[\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{t=1}^{T} \|X_{ijt}\|^{2q}_\infty] \Big)^{1/2} \Big(E[\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{t=1}^{T} |\varepsilon_{ijt}|^{2q}] \Big)^{1/2} \Big]^{2/q} \\ =& O((NM)^{2/q} B^2_{NM})\end{aligned}$$

where the first inequality is due to Jensen's inequality, the third inequality is due to Hölder's inequality, and the last equality is due to Assumption 3 (1) and (3). Thus $\frac{B_{NM} \log(p\vee(NM))}{(NM)^{1-1/q}} = O(\sqrt{\frac{\log a}{NM}})$, and this implies

$$2c\|X'\varepsilon\|_\infty = \max_{l\in[k_0]} \Big| \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{t=1}^{T} (X_{ijt,l}\varepsilon_{ijt} - E[X_{ijt,l}\varepsilon_{ijt}]) \Big| \lesssim \sqrt{NM \log a} = C^{-1}\mu$$

with probability at least $1 - C(\log(NM))^{-1}$ for $K > 0$ large enough.

Since $\|(D_1, D_2)\|_\infty = 1$ under Assumption 3 (2), an application of Lemma B.2 gives

$$2c\|D_1'\varepsilon\|_\infty = \max_{l\in\{k_0+1,\ldots,k_0+N_0\}} \Big|\sum_{j=1}^{M}\sum_{t=1}^{T}(D_{1,ijt,l} - ED_{1,ijt,l})\Big| \lesssim \sqrt{M\log a} = C^{-1}\mu/\sqrt{N}$$

with probability at least $1 - C(\log(N \wedge M))^{-1}$, where i depends on the choice of l. Note that there are at most $MT = O(M)$ nonzero terms in the summand for each l. Analogous arguments hold for $\|D_2'\varepsilon\|_\infty$ with the number of nonzero terms being at most $NT = O(N)$ in place of MT.

Under Assumption 3 and the choice $\mu^l_{\text{node}} = C\sqrt{NM\log a}$, similar lines of argument to those above show that the regularized events $\|X^{-l}\zeta^l\|_\infty \leq \mu^l_{\text{node}}/2c$, $\|D_1\zeta^l\|_\infty \leq \mu^l_{\text{node}}/2c\sqrt{N}$, and $\|D_2\zeta^l\|_\infty \leq \mu^l_{\text{node}}/2c\sqrt{M}$ occur with probability approaching one. Applying Lemma B.2, we have

$$\begin{aligned}\frac{\|Z^{-l}\zeta^l\|_\infty}{NM} &\leq \max_{k\in[k_0-1],l\in[p]}\Big|\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{t=1}^{T}(Z^{-l}_{ijt,k}\zeta^l_{ijt} - E[Z^{-l}_{ijt,k}\zeta^l_{ijt}])\Big| \\ &\lesssim \sqrt{\frac{\sigma^2\log(p^2\vee(NM))}{NM}} + \frac{B\log(p^2\vee(NM))}{NM} \lesssim \sqrt{\frac{\log a}{NM}}\end{aligned}$$

with probability $1 - C(\log(N \wedge M))^{-1}$. □

B.4 Rates of Nuisance Parameters

Throughout this section, we use the following notation. For any diagonal matrix A, A^l denotes the l-th diagonal entry and A^{-l} denotes A with the l-th column and row removed. Also, denote $\bar{\Psi}^{-l,-l}$ for $\bar{\Psi}$ with its l-th column and row removed.

The following lemma establishes the behavior of the nuisance parameters based on nodewise regressions under the three-dimensional panel setting. It is closely related to Lemma C.9 of Kock and Tang (2019). The main difference is that, in Kock and Tang (2019), their one-way fixed effect modeling assumption implies their $D_2 = \emptyset$ and $D_1'D_1 = I$, which in turn implies the diagonal structure of

$$\Theta = \begin{bmatrix}\Theta_X & 0 \\ 0 & I\end{bmatrix},$$

and greatly simplifies their estimation procedure. In our case, however, due to the potential presence of multi-way fixed effects, such a decomposition is not available. Therefore, the theory of our nodewise regression needs to simultaneously account for these fixed effects with different convergence rates.

Lemma B.4 (Nodewise Lasso for Nuisance Parameters). *Suppose Assumptions 3, 4, and 5 are satisfied and $\widehat{\Theta}$ is calculated following (3.3) with $\mu^l_{node} = C\sqrt{NM\log a}$ for a $C > 0$. It holds uniformly over $l \in [k_0]$ that*

$$\begin{aligned}
\|\widehat{\phi}^l - \phi^l\|_1 =& O_p\Big(\sqrt{\frac{s_l^2 \log a}{N \wedge M}}\Big), \\
\|\widehat{\phi}^l - \phi^l\| =& O_p\Big(\sqrt{\frac{s_l \log a}{N \wedge M}}\Big), \\
|\widehat{\tau}_l^2 - \tau_l^2| =& O_p\Big(\sqrt{\frac{s_l \log a}{NM}}\Big), \\
\Big|\frac{1}{\widehat{\tau}_l^2} - \frac{1}{\tau_l^2}\Big| =& O_p\Big(\sqrt{\frac{s_l \log a}{NM}}\Big), \\
\|\widehat{\Theta}_l' - \Theta_l'\|_1 =& O_p\Big(\sqrt{\frac{s_l^2 \log a}{N \wedge M}}\Big), \\
\|\widehat{\Theta}_l' - \Theta_l'\| =& O_p\Big(\sqrt{\frac{s_l \log a}{N \wedge M}}\Big), \\
\|\widehat{\Theta}_l\|_1 =& O_p(s_l^{1/2}), \qquad \text{and} \\
\max_{l\in[k_0]}\Big|\frac{1}{\widehat{\tau}_l^2}\Big| =& O_p(1).
\end{aligned}$$

Furthermore, these bounds are valid uniformly over the ℓ_0-ball $\{\overline{\eta} \in \mathbb{R}^p : \|\overline{\eta}\|_0 \le s_l\}$.

Proof. The proof consists of three steps.

Step 1 First, under Assumption 3 and by the choice $\mu^l_{\text{node}} = C\sqrt{NM\log a}$, Lemma B.3 gives that the regularized events $\|X^{-l}\zeta^l\|_\infty \le \mu^l_{\text{node}}/2c$, $\|D_l\zeta^l\|_\infty \le \mu^l_{\text{node}}/2c\sqrt{N}$, $\|D_2\zeta^l\|_\infty \le \mu^l_{\text{node}}/2c\sqrt{M}$ occur with probability approaching one uniformly over $[k_0]$. Under Assumptions 4 and 5 (1) and (2), apply Lemma B.1 with $(Z^{-l}, Z^l, \zeta^l, R^l, \phi^l)$ in place of $(Z, Y, \varepsilon, R, \overline{\boldsymbol{\eta}})$, we have the following oracle inequality

$$\|Z^{-l}(\widehat{\phi}^l - \phi^l)\| = \sqrt{(\widehat{\phi}^l - \phi^l)'Q^{-l}\bar{\Psi}^{-l,-l}Q^{-l}(\widehat{\phi}^l - \phi^l)} \lesssim \frac{\mu^l_{\text{node}}\sqrt{s_l}}{\sqrt{NM}\kappa^l_{c_0}} + c_{s_l},$$

where the restricted eigenvalue $\kappa^l_{c_0}$ is defined as in (B.1) with $\bar{\Psi}^{-l,-l}$ replacing $\bar{\Psi}$. Observe that $\kappa^l_{c_0} \ge \underline{k} > 0$ is satisfied with probability $1 - o(1)$ following Assumption 4. Hence by replacing μ^l_{node} in the oracle inequality with its upper bound, one has

$$\frac{1}{NM}\|Z^{-l}(\widehat{\phi}^l - \phi^l)\|^2 = (\widehat{\phi}^l - \phi^l)'Q^{-l}\bar{\Psi}^{-l,-l}Q^{-l}(\widehat{\phi}^l - \phi^l) = O_p\Big(\frac{s_l \log a}{NM}\Big). \tag{B.6}$$

Similarly, by Lemma B.1, under Assumptions 4 and 5 (1) and (2)

$$\|Q^{-l}(\widehat{\phi}^l - \phi^l)\|_1 = O_p\Big(s_l\sqrt{\frac{\log a}{NM}}\Big), \tag{B.7}$$

$$\|\widehat{\phi}^l - \phi^l\|_1 = O_p\Big(s_l\sqrt{\frac{\log a}{N \wedge M}}\Big), \tag{B.8}$$

uniformly for $l \in [k_0]$.

Now, to find a bound for $\|\widehat{\phi}^l - \phi^l\|$ that holds uniformly over $[k_0]$, note that

$$\begin{aligned}(\widehat{\phi}^l - \phi^l)' Q^{-l}\Psi^{-l,-l}Q^{-l}(\widehat{\phi}^l - \phi^l) \leq &(\widehat{\phi}^l - \phi^l)' Q^{-l}\bar{\Psi}^{-l,-l}Q^{-l}(\widehat{\phi}^l - \phi^l) + \|\bar{\Psi} - \Psi\|_\infty \|Q^{-l}(\widehat{\phi}^l - \phi^l)\|_1^2. \\ \leq &O_p\Big(\frac{s\log a}{NM}\Big) + \|\bar{\Psi} - \Psi\|_\infty \|Q^{-l}(\widehat{\phi}^l - \phi^l)\|_1^2, \end{aligned} \tag{B.9}$$

by (B.6), where $\|A\|_\infty$ denotes the maximal element of a matrix A. We now bound the second term on the right-hand side. Note that

$$\begin{aligned} P\Big(\|\bar{\Psi} - \Psi\|_\infty \geq r\Big) \leq &P\Big(\max_{t\in[T]}\max_{l\in[k_0]}\Big\|\frac{1}{NM}\sum_{i,j}(X^2_{ijt,l} - EX^2_{ijt,l})\Big\|_\infty \geq r/T\Big) \\ &+P\Big(\max_{t\in[T]}\max_{l\in\{k_0+1,\ldots,k_0+N_0\}}\Big\|\frac{1}{M}\sum_{j}(D^2_{1,ijt,l} - ED^2_{1,ijt,l})\Big\|_\infty \geq r/T\Big) \\ &+P\Big(\max_{t\in[T]}\max_{l\in\{k_0+N_0+1,\ldots,k_0+N_0+M_0\}}\Big\|\frac{1}{N}\sum_{i}(D^2_{2,ijt,l} - ED^2_{2,ijt,l})\Big\|_\infty \geq r/T\Big). \end{aligned}$$

We want to show all three terms go to zero with $r = C\sqrt{\frac{\log a}{NM}}$. Assumption 3 (1) and (3) imply

$$\begin{aligned} B^2 = E[\max_{i\leq N, j\leq M, t\leq T}\|X_{ijt}\|^4_\infty] \leq &(E[\max_{i\leq N, j\leq M, t\leq T}\|X_{ijt}\|^{2q}_\infty])^{2/q} \\ \leq &(NM)^{2/q}\Big(E\Big[\frac{1}{NM}\sum_{i=1}^N\sum_{j=1}^M\max_{t\leq T}\|X_{ijt}\|^{2q}_\infty\Big]\Big)^{2/q} \\ \leq &(NM)^{2/q}B^4_{NM}. \end{aligned}$$

Thus, with probability at least $1 - C(\log(NM))^{-1}$,

$$\max_{t\in[T]}\max_{l\in[k_0]}\Big\|\frac{1}{NM}\sum_{i,j}(X^2_{ijt,l} - EX^2_{ijt,l})\Big\|_\infty \lesssim \sqrt{\frac{\log(k_0^2 \vee (NM))}{NM}} + \frac{B^2_{NM}\log(k_0^2 \vee (NM))}{(NM)^{1-1/q}} \lesssim \sqrt{\frac{\log a}{NM}},$$

by Lemma B.2. Similarly, with probability at least $1-C((\log N\wedge M))^{-1}$,

$$\max_{t\in[T]}\max_{l\in\{k_0+1,\dots,k_0+N_0\}}\Big\|\frac{1}{M}\sum_j(D^2_{1,ijt,l}-ED^2_{1,ijt,l})\Big\|_\infty\lesssim\sqrt{\frac{\log M}{M}}+\frac{\log M}{M}\lesssim\sqrt{\frac{\log a}{M}},$$
$$\max_{t\in[T]}\max_{l\in\{k_0+N_0+1,\dots,k_0+N_0+M_0\}}\Big\|\frac{1}{N}\sum_i(D^2_{2,ijt,l}-ED^2_{2,ijt,l})\Big\|_\infty\lesssim\sqrt{\frac{\log N}{N}}+\frac{\log N}{N}\lesssim\sqrt{\frac{\log a}{N}}$$

by Assumption 3 (2). Thus, with probability at least $1-C(\log(N\wedge M))^{-1}$,

$$\|\bar{\Psi}-\Psi\|_\infty=O_p\Big(\sqrt{\frac{\log a}{N\wedge M}}\Big).$$

Following Assumption 5(4), $s_l\sqrt{\frac{\log a}{N\wedge M}}=o(1)$, we therefore have

$$\begin{aligned}\|\bar{\Psi}-\Psi\|_\infty\|Q^{-l}(\widehat{\phi}^l-\phi^l)\|_1^2=&O_p\Big(\sqrt{\frac{\log a}{N\wedge M}}\Big)O_p\Big(\frac{s_l^2\log a}{NM}\Big)\\=&O_p\Big(s_l\sqrt{\frac{\log a}{N\wedge M}}\Big)O_p\Big(\frac{s_l\log a}{NM}\Big)=o_p\Big(\frac{s_l\log a}{NM}\Big)\end{aligned}$$

uniformly in $l\in[k_0]$. Substituting into (B.9), we obtain

$$(\widehat{\phi}^l-\phi^l)'Q^{-l}\Psi^{-l,-l}Q^{-l}(\widehat{\phi}^l-\phi^l)=O_p\Big(\frac{s\log a}{NM}\Big)$$

uniformly in $l\in[k_0]$. Since under Assumption 5(3), $\Lambda_{\min}(\Psi)>0$ and

$$\Lambda_{\min}(\Psi)\|\widehat{Q}^{-l}(\phi^l-\phi^l)\|^2\le\max_{l\in[k_0]}(\widehat{\phi}^l-\phi^l)'Q^{-l}\Psi^{-l,-l}Q^{-l}(\widehat{\phi}^l-\phi^l)$$

uniformly in $l\in[k_0]$, we conclude that

$$\|Q^{-l}(\widehat{\phi}^l-\phi^l)\|=O_p\Big(\sqrt{\frac{s_l\log a}{NM}}\Big),$$

uniformly in $l\in[k_0]$. Thus, the triangle inequality and definition of Q together imply it holds uniformly over $[k_0]$

$$\|\widehat{\phi}^l-\phi^l\|=O_p\Big(\sqrt{\frac{s_l\log a}{N\wedge M}}\Big).$$

Step 2 We next derive the rate of $\max_{l\in[k_0]}|\widehat{\tau}_l^2-\tau_l^2|$. By the definition of $\widehat{\tau}_l$ and the K.K.T. condition, we get the following equality using the decomposition $Z^l = Z^{-l}\phi^l + R^l + \zeta^l$.

$$\begin{aligned}
\widehat{\tau}_l^2 =& \frac{(Z^l - Z^{-l}\widehat{\phi}^l)'Z^l}{NM} \\
=& \frac{[R^l + \zeta^l - Z^{-l}(\widehat{\phi}^l - \phi^l)]'(Z^{-l}\phi^l + r^l + \zeta^l)}{NM} \\
=& \frac{(\zeta^l)'\zeta^l}{NM} + \frac{(\zeta^l)'Z^{-l}\phi^l}{NM} - \frac{(\widehat{\phi}^l - \phi^l)'(Z^{-l})'Z^{-l}\phi^l}{NM} - \frac{(\widehat{\phi}^l - \phi^l)'(Z^{-l})'\zeta^l}{NM} \\
&+ \Big(\frac{(R^l)'R^l}{NM} + \frac{(R^l)'Z^{-l}\phi^l}{NM} + \frac{2(R^l)'\zeta^l}{NM} - \frac{(\widehat{\phi}^l - \phi^l)'Z^{-l}R^l}{NM}\Big).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\max_{l\in[k_0]} |\widehat{\tau}_l^2 - \tau_l^2| \\
\leq& \max_{l\in[k_0]} \Big|\frac{(\zeta^l)'\zeta^l}{NM} - \tau_l^2\Big| + \max_{l\in[k_0]} \Big|\frac{(\zeta^l)'Z^{-l}\phi^l}{NM}\Big| + \max_{l\in[k_0]} \Big|\frac{(\widehat{\phi}^l - \phi^l)'(Z^{-l})'Z^{-l}\phi^l}{NM}\Big| + \max_{l\in[k_0]} \Big|\frac{(\widehat{\phi}^l - \phi^l)'(Z^{-l})'\zeta^l}{NM}\Big| \\
&+ \max_{l\in[k_0]} \Big|\frac{(R^l)'R^l}{NM} + \frac{(R^l)'Z^{-l}\phi^l}{NM} + \frac{2(R^l)'\zeta^l}{NM} - \frac{(\widehat{\phi}^l - \phi^l)'Z^{-l}R^l}{NM}\Big| = (i) + (ii) + (iii) + (iv) + (v).
\end{aligned} \tag{B.10}$$

It suffices to find bounds for each of the five terms in the last expression.

First we consider (i). Under Assumption 3 (1), we have

$$\begin{aligned}
E\Big[\max_{i,j,t} |\zeta_{ijt}^l|^4\Big] =& \Big(E\Big[\max_{i,j,t} |\zeta_{ijt}^l|^{2q}\Big]\Big)^{2/q} \\
=& (NM)^{2/q}\Big(E\Big[\frac{1}{NM}\sum_{i,j} \max_t |\zeta_{ijt}^l|^{2q}\Big]\Big)^{2/q} \\
\lesssim& (NM)^{2/q}\Big(E\Big[\frac{1}{NM}\sum_{i=1}^N\sum_{j=1}^M \max_{t\leq T} \|X_{ijt}\|_\infty^{2q}\Big]\Big)^{2/q} \leq (NM)^{2/q} B_{NM}^4
\end{aligned}$$

for all $l \in [k_0]$. Therefore, by Lemma B.2 and Assumption 3 (1),

$$\begin{aligned}
\max_{l\in[k_0]} \Big\|\frac{1}{NM}\sum_{i,j,t}[(\zeta_{ijt}^l)^2 - E(\zeta_{ijt}^l)^2]\Big\|_\infty \leq& T \max_{t\in[T]}\max_{l\in[k_0]} \Big\|\frac{1}{NM}\sum_{i,j}[(\zeta_{ijt}^l)^2 - E(\zeta_{ijt}^l)^2]\Big\|_\infty \\
\lesssim& \sqrt{\frac{\log(k_0 \vee (NM))}{NM}} + \frac{B_{NM}^2 \log(k_0 \vee (NM))}{(NM)^{1-1/q}} \lesssim \sqrt{\frac{\log a}{NM}}
\end{aligned}$$

with probability at least $1 - C(\log(NM))^{-1}$. It follows that

$$\max_{l\in[k_0]} \Big|\frac{(\zeta^l)'\zeta^l}{NM} - \tau_l^2\Big| = O_p\Big(\sqrt{\frac{\log a}{NM}}\Big).$$

Second, we consider (iv) in (B.10). By the regularized events established in Step 1, we have

$$\Big\|\frac{(Q^{-l})^{-1}(Z^{-l})'\zeta^l}{NM}\Big\|_\infty = \max\Big\{\Big\|\frac{(X^{-l})'\zeta^l}{NM}\Big\|_\infty, \sqrt{N}\Big\|\frac{D_1'\zeta^l}{NM}\Big\|_\infty, \sqrt{M}\Big\|\frac{D_2'\zeta^l}{NM}\Big\|_\infty\Big\} = O_p\Big(\sqrt{\frac{\log a}{NM}}\Big).$$

Thus, by (B.7),

$$\max_{l\in[k_0]}\Big|\frac{(\widehat{\phi}^l-\phi^l)'(Z^{-l})'\zeta^l}{NM}\Big| \leq \Big\|\frac{(Q^{-l})^{-1}(Z^{-l})'\zeta^l}{NM}\Big\|_\infty \|Q^{-l}(\widehat{\phi}^l-\phi^l)\|_1 = O_p\Big(\frac{s_l\log p}{NM}\Big) = O_p\Big(\sqrt{\frac{s_l\log p}{NM}}\Big)$$

follows.

We next consider (ii) in (B.10). Note that $\|\phi^l\|_1 = O(\sqrt{s_l})$ in Assumption 5(1) implies $\|Q^{-l}\phi^l\|_1 = O(\sqrt{s_l})$. Combining this and the regularized events as before, we have

$$\max_{l\in[k_0]}\Big|\frac{(\zeta^l)'Z^{-l}\phi^l}{NM}\Big| \leq \max_{l\in[k_0]}\Big\|\frac{(\zeta^l)'Z^{-l}(Q^{-l})^{-1}}{NM}\Big\|_\infty\|Q^{-l}\phi^l\|_1 = O_p\Big(\sqrt{\frac{s_l\log a}{NM}}\Big).$$

Now, we consider (iii) in (B.10). Using Assumptions 4 and 5 (1) and (3), the fact that $Q^{-1} = \sqrt{NM}S^{-1}$, and the definition of $\bar{\Psi}$, we obtain

$$\begin{aligned}
\|Z^{-l}\phi^l\| \leq& \|Q^{-l}\phi^l\| \max_{\substack{\|\xi\|=1\\ \|\xi\|_0\leq s_l}} \sqrt{\xi'(Q^{-l})^{-1}(Z^{-l})'Z^{-l}(Q^{-l})^{-1}\xi}\\
\leq& \sqrt{NM}\|Q^{-l}\phi^l\| \max_{\substack{\|\xi\|=1\\ \|\xi\|_0\leq s_l}} \sqrt{\xi'\bar{\Psi}\xi}\\
\leq& \sqrt{NM}\cdot O(1)\cdot\sqrt{\varphi_{\max}(\bar{\Psi},s_l)} = O_p(\sqrt{NM}).
\end{aligned}$$

Furthermore, (B.6) implies,

$$\frac{1}{NM}\|Z^{-l}(\widehat{\phi}^l-\phi^l)\| \leq O_p\Big(\frac{(s_l\log a)^{1/2}}{NM}\Big).$$

Combining these two intermediate results, we have

$$\max_{l\in[k_0]}\Big|\frac{(\widehat{\phi}^l-\phi^l)'(Z^{-l})'Z^{-l}\phi^l}{NM}\Big| \leq \max_{l\in[k_0]}\frac{\|Z^{-l}(\widehat{\phi}^l-\phi^l)\|\|Z^{-l}\phi^l\|}{NM} = O_p\Big(\sqrt{\frac{s_l\log a}{NM}}\Big).$$

Finally, we consider the remaining terms in (B.10) that involve r_l. Note that

$$\begin{aligned}
\frac{|(R_l)'R_l|}{NM} \lesssim& \frac{s_l}{NM},\\
\frac{|(R_l)'\zeta^l|}{NM} \leq& \frac{1}{NM}\|R^l\|\|\zeta^l\| \leq \frac{1}{\sqrt{NM}}\sqrt{s_l}O_p\Big(\sqrt{\max_{t\in[T]}\max_{l\in[k_0]}\frac{T}{NM}\sum_{i=1}^N\sum_{j=1}^M E(\zeta_{ijt}^l)^2}\Big) \lesssim O_p\Big(\sqrt{\frac{s_l}{NM}}\Big)
\end{aligned}$$

follows from Assumption 5 (2) and Assumption 3 (3). Also, under Assumptions 4 and 5 (1) and (2)

$$\begin{aligned}\frac{|(R^l)'Z^l\phi^l|}{NM} \leq& \frac{\|R^l\|\|Z^{-l}\phi^l\|}{NM} \\ \leq& \frac{1}{NM}\sqrt{s_l}\|Q^{-l}\phi^l\| \max_{\substack{\|\delta\|=1 \\ \|\delta\|_0\leq s_l}} \sqrt{\delta'(Q^{-l})^{-1}(Z^{-l})'Z^{-l}(Q^{-l})^{-1}\delta} \\ \leq& \frac{1}{\sqrt{NM}}\sqrt{s_l}O(1) \max_{\substack{\|\delta\|=1 \\ \|\delta\|_0\leq s_l}} \sqrt{\delta'\bar{\Psi}\delta} \\ \leq& \sqrt{\frac{s_l}{NM}}\sqrt{\varphi_{\max}(\bar{\Psi}, s_l)} = O\Big(\sqrt{\frac{s_l}{NM}}\Big)\end{aligned}$$

with probability at least $1-o(1)$. A similar argument under Assumptions 4 and 5 (1) and (2) shows that $\frac{|(R^l)'Z^l(\widehat{\phi}^l-\phi^l)|}{NM} = O\Big(\sqrt{\frac{s_l}{NM}}\Big)$.

Combining all the results above, we obtain

$$|\widehat{\tau}_l^2 - \tau_l^2| = O_p\Big(\sqrt{\frac{s_l \log a}{NM}}\Big)$$

uniformly over $[k_0]$.

Step 3 Since $l \in [k_0]$,

$$\frac{1}{\tau_l^2} = \Theta_{l,l} = (Q^l)^{-1}\Theta_{l,l}(Q^l)^{-1} \leq \Lambda_{\max}(Q^{-1}\Theta Q^{-1}) = \Lambda_{\max}(\Psi^{-1}) = 1/\Lambda_{\min}(\Psi) = O(1), \quad \text{(B.11)}$$

holds for each (N, M) under Assumption 5 (3), where the first inequality follows from the discussion following (B.30) in the Proof of Theorem 1 in Caner and Kock (2018). Therefore, $\widehat{\tau}_l^2$ is bounded away from zero in probability, and we have

$$\max_{l\in[k_0]}\Big|\frac{1}{\widehat{\tau}_l^2} - \frac{1}{\tau_l^2}\Big| = O_p\Big(\sqrt{\frac{s_l \log a}{NM}}\Big) \quad \text{(B.12)}$$

by Step 2.

Now, we bound $\max_{l\in[k_0]}\|\widehat{\Theta}_l - \Theta_l\|_1$. Since $\|\phi^l\|_1 = O(\sqrt{s_l})$ under Assumption 5 (1), we have

$$\begin{aligned}\max_{l\in[k_0]}\|\widehat{\Theta}_l - \Theta_l\|_1 \leq& \max_{l\in[k_0]}\Big\|\frac{\widehat{C}_l}{\widehat{\tau}_l^2} - \frac{C_l}{\tau_l^2}\Big\|_1 \\ \leq& \max_{l\in[k_0]}\Big|\frac{1}{\widehat{\tau}_l^2} - \frac{1}{\tau_l^2}\Big| + \max_{l\in[k_0]}\Big\|\frac{\widehat{\phi}^l}{\widehat{\tau}_l^2} - \frac{\phi^l}{\widehat{\tau}_l^2} + \frac{\phi^l}{\widehat{\tau}_l^2} - \frac{\phi^l}{\tau_l^2}\Big\|_1 \\ \leq& \max_{l\in[k_0]}\Big|\frac{1}{\widehat{\tau}_l^2} - \frac{1}{\tau_l^2}\Big| + \max_{l\in[k_0]}\frac{\|\widehat{\phi}^l - \phi^l\|_1}{\widehat{\tau}_l^2} + \max_{l\in[k_0]}\|\phi^l\|_1 \max_{l\in[k_0]}\Big(\Big|\frac{1}{\widehat{\tau}_l^2} - \frac{1}{\tau_l^2}\Big|\Big).\end{aligned}$$

The first and the third terms can be bounded by (B.12) and the second term can be bounded by (B.8). Therefore,

$$\begin{aligned}\max_{l\in[k_0]}\|\widehat{\Theta}_l-\Theta_l\|_1 =&O_p\Big(\frac{s_l\log a}{NM}\Big)+O_p\Big(s_l\sqrt{\frac{\log a}{N\wedge M}}\Big)+O_p\Big(s_l\sqrt{\frac{\log a}{NM}}\Big)\\ =&O_p\Big(s_l\sqrt{\frac{\log a}{N\wedge M}}\Big).\end{aligned}$$

Similar lines of argument under Assumption 5 (1) and $\|\widehat{\phi}^l-\phi^l\|$ from Step 1 lead to

$$\|\widehat{\Theta}_l-\Theta_l\| =O_p\Big(\sqrt{\frac{s_l\log a}{N\wedge M}}\Big).$$

Since $\|\Theta_l\|_1\le\max_{l\in[k_0]}\frac{1}{\tau_l^2}+\max_{l\in[k_0]}\|\frac{\phi^l}{\tau_l^2}\|_1=O(\sqrt{s_l})$ by (B.11) and Assumption 5 (1), it follows that $\|\widehat{\Theta}_l\|_1=O_p(\sqrt{s_l})$ for all $l\in[k_0]$.

□

B.5 Sufficiency for Assumption 1 (i)

Lemma B.5. *If Assumptions 2, 3, 4, and 5 are satisfied, then*

$$\max_{l\in[k_0]}\Big|\sqrt{NM}(\widehat{\Theta}_l'Q\bar{\Psi}Q-e_l')(\widehat{\boldsymbol{\eta}}-\overline{\boldsymbol{\eta}})\Big|=o_p(1).$$

Proof. Recall $\bar{\Psi}=S^{-1}Z'ZS^{-1}$ and $Q=S/\sqrt{NM}$. Also, if we let $\Gamma=ZS^{-1}$, then $\bar{\Psi}=\Gamma'\Gamma$ and

$$\frac{Z'Z}{NM}=Q\Gamma'\Gamma Q=Q\bar{\Psi}Q.$$

Since $l\in[k_0]$, $Q_{ll}=1$. Let $\bar{\Psi}_l$, Γ_l denote the l-th column of $\bar{\Psi}$ and Γ and $\bar{\Psi}_{-l}$, Γ_{-l} be the respective original matrices with l-th column removed. Using the K.K.T. condition for the nodewise lasso, we have

$$1=\frac{(Z^l-Z^{-l}\widehat{\phi}^l)'Z^l}{\widehat{\tau}_l^2NM}=\frac{\widehat{\Theta}_l'Z'Z^l}{NM}=\widehat{\Theta}_l'Q\Gamma'\Gamma_l\cdot 1=\widehat{\Theta}_l'Q\bar{\Psi}_lQ_{ll}. \tag{B.13}$$

Also using the K.K.T. condition, we have

$$\frac{Q^{-l}\widehat{\kappa}_l}{NM}=\frac{(Z^{-l})'(Z^l-Z^{-l}\widehat{\phi}^l)}{\mu^l_{\text{node}}NM}.$$

Using the property of the sub-gradient κ_l, we have

$$\Big\|\frac{(Z^{-l})'(Z^l - Z^{-l}\widehat{\phi}^l)}{\mu^l_{\text{node}} NM}\Big\|_\infty \le \frac{\|\widehat{\kappa}_l\|_\infty}{NM} \le \frac{1}{NM},$$

which is the same as

$$\Big\|\frac{(Z^{-l})'Z\widehat{C}_l}{NM}\Big\|_\infty \le \frac{\mu^l_{\text{node}}}{NM},$$

since $Z^l - Z^{-l}\widehat{\phi}^l = Z\widehat{C}_l$. Divide both sides by $\widehat{\tau}_l^2$ and use $\widehat{\Theta}_l = \widehat{C}_l/\widehat{\tau}_l^2$ to obtain

$$\Big\|\frac{(Z^{-l})'Z\widehat{\Theta}_l}{NM}\Big\|_\infty \le \frac{\mu^l_{\text{node}}}{\widehat{\tau}_l^2 NM}.$$

With some rewriting

$$\frac{\mu^l_{\text{node}}}{\widehat{\tau}_l^2 NM} \ge \Big\|\frac{(Z^{-l})'Z\widehat{\Theta}_l}{NM}\Big\|_\infty = \Big\|\frac{S^{-l}\Gamma'_{-l}\Gamma S\widehat{\Theta}_l}{NM}\Big\|_\infty = \Big\|\frac{S^{-l}\bar{\Psi}_{-l}S\widehat{\Theta}_l}{NM}\Big\|_\infty = \Big\|Q^{-l}\bar{\Psi}_{-l}Q\widehat{\Theta}_l\Big\|_\infty, \tag{B.14}$$

where S^{-l} is S with both the l-th column and the l-th row removed. Q^{-l} is defined similarly. Applying Lemma B.4 under Assumptions 3, 4, and 5, we have $1/\widehat{\tau}_l^2 = O_p(1)$. Therefore, by (B.13) and (B.14),

$$\max_{l\in[k_0]}\Big\|\widehat{\Theta}_l'Q\bar{\Psi}Q - e_l'\Big\|_\infty = \max_{l\in[k_0]}\Big\|\frac{(Z^l)'X\widehat{\Theta}_l}{NM}\Big\|_\infty \lesssim \max_{l\in[k_0]}\frac{\mu_{\text{node}}}{\widehat{\tau}_l^2 NM} = O_p\Big(\sqrt{\frac{\log a}{NM}}\Big).$$

Finally, Lemma B.1 and Lemma B.3 with $\mu = C\sqrt{(NM)\log a}$ under Assumptions 2, 3, and 4 together imply

$$\begin{aligned}
&\max_{l\in[p]}|\sqrt{NM}(\widehat{\Theta}_l Q\bar{\Psi}Q - e_l')(\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}})| \le \sqrt{NM}\max_{l\in[k_0]}\Big\|\widehat{\Theta}_l'Q\bar{\Psi}Q - e_l'\Big\|_\infty\|\widehat{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}\|_1 \\
=&\sqrt{NM}O_p\Big(\sqrt{\frac{\log a}{NM}}\Big)O_p\Big(s\sqrt{\frac{\log a}{N\wedge M}}\Big) = O_p\Big(\sqrt{\frac{s^2(\log a)^2}{N\wedge M}}\Big) = o_p(1)
\end{aligned}$$

as claimed.[1] □

B.6 Sufficiency for Assumption 1 (ii)

Lemma B.6. *Suppose that Assumptions 2, 3, 4, and 5 are satisfied. Then,*

$$\max_{l\in[k_0]}\Big|\widehat{\Theta}_l'Z'R/\sqrt{NM}\Big| = o_p(1).$$

[1]Note that Lemma B.1, as it is stated, requires Assumptions 2, B.1, B.2, and B.3. While Assumption 2 is directly invoked by the statement of Lemma B.5, Assumption B.1 is implied by Assumption 3 through Lemma B.3, Assumption B.2 is trivially satisfied under the current setting with $\widehat{\Upsilon} = I$, and Assumption B.3 is implied by Assumption 4.

Proof. Note that

$$\max_{l\in[k_0]}\|\widehat{\Theta}_l\| = O_p(1)$$

by Assumption 5 (1) and (4) and Lemma B.4 under Assumptions 3, 4, and 5. Therefore,

$$\max_{l\in[k_0]}\left|\widehat{\Theta}_l' Z' R/\sqrt{NM}\right| \le \frac{1}{NM}\max_{l\in[k_0]}\|\widehat{\Theta}_l\|\|Z'R\| = o_p(1)$$

follows under Assumption 2 (3). □

B.7 Sufficiency for Assumption 1 (iii)

Lemma B.7. *Suppose that Assumptions 3, 4, 5 and 6 are satisfied. Then,*

$$V_{ll}^{-1/2}\widehat{\Theta}_l' Z'\varepsilon/\sqrt{NM} \rightsquigarrow N(0,1).$$

Proof. First we show $\frac{1}{\sqrt{NM}}\Theta_l' Z'\varepsilon \rightsquigarrow N(0,V_{ll})$. Note that we have

$$E\left[\frac{1}{\sqrt{NM}}\Theta_l' Z'\varepsilon\right] = \frac{1}{\sqrt{NM}}E[\Theta_l' Z' E[\varepsilon|Z]] = 0$$

and

$$V_{ll} = E\left[\Big(\frac{1}{\sqrt{NM}}\Theta_l' Z'\varepsilon\Big)\Big(\frac{1}{\sqrt{NM}}\Theta_l' Z'\varepsilon\Big)'\right] = \Theta_l'\Omega\Theta_l \ge \underline{k} > 0$$

under Assumption 6. Furthermore, by Assumption 3

$$\begin{aligned}
E\left|\frac{1}{\sqrt{NM}}\Theta_l' Z'\varepsilon\right|^q \le& \frac{1}{(NM)^{q/2}}E\|\Theta_l\|_1^q \max_{k\in\text{supp}(\Theta_l)}\sum_{i=1}^N\sum_{j=1}^M\sum_{t=1}^T\left|Z_{ijt,k}\varepsilon_{ijt}\right|^q \\
\lesssim& \frac{s_l^{q/2}}{(NM)^{q/2}}\sum_{k\in\text{supp}(\Theta_l)} E\sum_{i=1}^N\sum_{j=1}^M\sum_{t=1}^T\left|Z_{ijt,k}\varepsilon_{ijt}\right|^q \\
\le& \frac{s_l^{q/2}\cdot s_l}{(NM)^{q/2}}\max_{l\in[p]} E\sum_{i=1}^N\sum_{j=1}^M\sum_{t=1}^T\left|Z_{ijt,k}\varepsilon_{ijt}\right|^q \\
\le& \frac{s_l^{q/2+1}(NM)}{(NM)^{q/2}}\max_{l\in[p]}\sqrt{\frac{1}{NM}E\sum_{i=1}^N\sum_{j=1}^M\sum_{t=1}^T\left|Z_{ijt,k}\right|^{2q}\frac{1}{NM}E\sum_{i=1}^N\sum_{j=1}^M\sum_{t=1}^T\left|\varepsilon_{ijt}\right|^{2q}} \\
\le& \frac{s_l^{q/2+1}}{(NM)^{q/2-1}}O(1) = o(1),
\end{aligned}$$

where $q > 4$, the first inequality follows from a dual norm inequality, the second and the third from the fact that $\|\Theta_l\|_1 \lesssim \sqrt{s_l}$ and $\|\Theta_l\|_0 \leq s_l$ implied by Assumption 5(1), the fourth from Cauchy-Schwartz's inequality, and the fifth from Assumption 3 and the last equality follows from Assumption 5 (4). This verifies the Lyapunov's condition. Thus, we have $\frac{1}{\sqrt{NM}}\Theta_l' Z'\varepsilon \rightsquigarrow N(0, V_{ll})$.

Now, we show $|\frac{1}{\sqrt{NM}}(\widehat{\Theta}_l - \Theta_l)' Z'\varepsilon| = o_p(1)$. Invoking Lemmata B.3 and B.4 under Assumptions 3, 4, 5, we have

$$\begin{aligned}\left|\frac{1}{\sqrt{NM}}(\widehat{\Theta}_l - \Theta_l)' Z'\varepsilon\right| \leq & \|\widehat{\Theta}_l - \Theta_l\|_1 \left\|\frac{1}{\sqrt{NM}} Z'\varepsilon\right\|_\infty \\ \leq & O_p\Big(\sqrt{\frac{s_l^2 \log a}{N \wedge M}}\Big) O_p\Big(\sqrt{\log a}\Big) \\ = & O_p\Big(\sqrt{\frac{s_l^2 (\log a)^2}{N \wedge M}}\Big) = o_p(1).\end{aligned}$$

Combining these results concludes $\frac{1}{\sqrt{NM}}\widehat{\Theta}_l' Z'\varepsilon \rightsquigarrow N(0, V_{ll})$. □

B.8 Empirical Pre-Sparsity

The following lemma is a minor modification of Lemma 8 in Belloni et al. (2012).

Lemma B.8 (Empirical Pre-sparsity). *If Assumptions 2, 3, 4, and 5 are satisfied, then we have*

$$\widehat{s}_\ell = O_p(s_\ell) \text{ and } \widehat{s} = O_p(s),$$

where $\widehat{s}_\ell = \|\widehat{\phi}^\ell\|_0$ *and* $\widehat{s} = \|\widehat{\boldsymbol{\eta}}\|_0$.

Proof. Let $\hat{m}_\ell = \left|\hat{T}_\ell \setminus T_\ell\right|$, where $T_\ell = \text{supp}(\phi^\ell)$ and $\hat{T}_\ell = \text{supp}(\hat{\phi}^\ell)$. From K.K.T. condition, we have

$$2((Q^{-\ell})^{-1}(Z^{-\ell})'(Z^\ell - Z^{-\ell}\hat{\phi}^\ell))_k = \mu^\ell_{\text{node}} \cdot \text{sign}(\hat{\phi}^\ell_k)$$

for all $\ell \in [k_0]$ and $k \in \hat{T}_\ell \setminus T_\ell$. Thus,

$$\begin{aligned}\mu^\ell_{\text{node}}\sqrt{\hat{m}_\ell} \leq & 2\|((Q^{-\ell})^{-1}(Z^{-\ell})^{-1}(Z^\ell - Z^{-\ell}\hat{\phi}^\ell))_{\hat{T}_\ell \setminus T_\ell}\| + 2\|((Q^{-\ell})^{-1}(Z^{-\ell})' R^\ell)_{\hat{T}_\ell \setminus T_\ell}\| \\ & + 2\|((Q^{-\ell})^{-1}(Z^{-\ell})' Z^{-\ell}(\hat{\phi}^\ell - \phi^\ell))_{\hat{T}_\ell \setminus T_\ell}\| \\ = & (1) + (2) + (3). \qquad \text{(B.15)}\end{aligned}$$

We bound the three terms in the last expression separately. First, Lemma B.3 under Assumption 3 yields

$$(1) \leq 2\sqrt{\hat{m}_\ell}\|(Q^{-\ell})^{-1}(Z^\ell)'\zeta^\ell\|_\infty \leq \sqrt{\hat{m}_\ell}\frac{\mu^\ell_{\text{node}}}{c}$$

with probability at least $1 - C(\log(N \wedge M))^{-1}$. Second,

$$\begin{aligned}
\|((Q^{-\ell})^{-1}(Z^{-\ell})'R^\ell)_{\hat{T}_\ell \setminus T_\ell}\| =& \sup_{\substack{\|\delta\|=1 \\ \|\delta\|_0 \leq \hat{m}_\ell}} |\delta'(Q^{-\ell})^{-1}(Z^{-\ell})'R^\ell| \\
\leq& \sup_{\substack{\|\delta\|=1 \\ \|\delta\|_0 \leq \hat{m}_\ell}} \|\delta'(Q^{-\ell})^{-1}(Z^{-\ell})'\|\|R^\ell\| \\
\leq& \sup_{\substack{\|\delta\|=1 \\ \|\delta\|_0 \leq \hat{m}_\ell}} (NM)\sqrt{\delta'\bar{\Psi}\delta}\sqrt{\frac{s_\ell}{NM}} \\
\leq& (NM)\sqrt{\varphi_{\max}(\bar{\Psi}, \hat{m}_\ell)}\sqrt{\frac{s_\ell}{NM}}
\end{aligned}$$

follows by Assumptions 4 and 5. Therefore,

$$(2) \leq 2(NM)\sqrt{\varphi_{\max}(\bar{\Psi}, \hat{m}_\ell)}\sqrt{\frac{s_\ell}{NM}}.$$

Finally, by Lemma B.4 under Assumptions 3, 4, 5, we obtain

$$\begin{aligned}
\|((Q^{-\ell})^{-1}(Z^{-\ell})'Z^{-\ell}(\hat{\phi}^\ell - \phi^\ell))_{\hat{T}_\ell \setminus T_\ell}\| \leq& \sup_{\substack{\|\delta\|=1 \\ \|\delta\|_0 \leq \hat{m}_\ell}} |\delta'(Q^{-\ell})^{-1}(Z^{-\ell})'Z^{-\ell}(\hat{\phi}^\ell - \phi^\ell)| \\
\leq& \sup_{\substack{\|\delta\|=1 \\ \|\delta\|_0 \leq \hat{m}_\ell}} \|\delta'(Q^{-\ell})^{-1}(Z^{-\ell})'\|\|Z^{-\ell}(\hat{\phi}^\ell - \phi^\ell)\| \\
\leq& (NM)\sqrt{\varphi_{\max}(\bar{\Psi}, \hat{m}_\ell)}\sqrt{\frac{s_\ell \log a}{NM}}
\end{aligned}$$

with probability at least $1 - C(\log(N \wedge M))^{-1}$, where the last inequality is due to Assumption 4 and Lemma B.4. Using these bounds and (B.15), we obtain

$$\sqrt{\hat{m}_\ell} \lesssim \sqrt{\varphi_{\max}(\bar{\Psi}, \hat{m}_\ell)}\sqrt{s_\ell} = O(\sqrt{s_\ell})$$

with probability $1 - o(1)$. Under Assumptions 2, 3, 4, and 5, the result for $\hat{s}$ can be established following analogous arguments. □

C Additional Discussions on Simulation Results

In this section, we present omitted details of the discussions on simulation results presented in Section 6.2 in the main text.

In the middle panel of Table 1, where the true data generating model is Model (II), OLS and FE-I are biased while FE-II and FE-III yield little biases. These results are consistent with the current simulation setting as OLS and FE-I mis-specify the true model while FE-II and FE-III correctly specify the true model. The bias of POST (2.1) is slightly larger than those of FE-II and FE-III, but much smaller than those of OLS and FE-I. In other words, POST (2.1) is de-biased to a large extent but not to the full extent so that desired balance between the bias and variance is maintained. FE-II, as the oracle estimator, yields a smaller root mean square error than OLS, FE-I, or FE-III. Furthermore, POST (2.1) yields an even smaller root mean square error than the oracle estimator, FE-II. The coverage frequency of FE-II, as the oracle estimator, is closer to the nominal level 95% than those of OLS, FE-I, or FE-III. POST (2.1) yields the coverage frequency as close to the nominal level as the oracle estimator, FE-II.

In the bottom panel of Table 1, where the true data generating model is Model (III), OLS, FE-I, and FE-II are biased while FE-III yields little bias. These results are consistent with the current simulation setting as OLS, FE-I, and FE-II mis-specify the true model while FE-III correctly specifies the true model. The bias of POST (2.1) is between those of OLS, FE-I, and FE-II and that of FE-III. As above, POST (2.1) balances the trade-off between bias and variance in the sense that it is de-biased to some extent, but not fully. POST (2.1) yields a smaller root mean square error than any other estimator, including the oracle estimator, FE-III. POST (2.1) also yields a coverage frequency closer to the nominal level than any estimator, including the oracle estimator, FE-III. In summary, we observe that, when the true model is rich, POST (2.1) is more precise than parsimonious estimators and allows for as accurate inference as the oracle estimator.

D Additional Simulations

This section presents additional simulation designs and simulation results based on them that we omitted from the main draft.

D.1 Alternative Sample Sizes

While the main text presents simulation results in Table 1 under the baseline design only for the sample size $N = 200$ ($NMT = 1900$), we now present results under alternative sample sizes. Tables D.1 and D.2 display Monte Carlo simulation results with the smaller sample sizes $N = 10$ ($NMT = 450$) and $N = 15$ ($NMT = 1050$), respectively. The overall patterns of these results resemble that of Table 1 presented in the main text, although the latter is more precise due to the larger sample size.

D.2 Alternative Simulation Designs

In the baseline model presented in Section 6 in the main text, the i and j fixed effects are generated by $\alpha_i = s_\alpha \big/ \big(i \cdot (\log(i+1))^{3/2}\big)$ and $\gamma_j = s_\gamma \big/ \big(j \cdot (\log(j+1))^{3/2}\big)$, where $s_\alpha = s_\gamma = 1$. We used this design because it satisfies our assumptions – see Examples 1 and 2 – and the proposed method therefore is expected to work well. In this appendix, we examine alternative fixed effect designs including those that are known to violate our assumptions.

D.2.1 Alternative FE Design (1)

The i and j fixed effects are generated by

$$\alpha_i \sim N\left(m_\alpha, s_\alpha^2 \Big/ \left(\sqrt{i} \cdot (\log(i+1))^3\right)\right)$$

and

$$\gamma_j \sim N\left(m_\gamma, s_\gamma^2 \Big/ \left(\sqrt{j} \cdot (\log(j+1))^3\right)\right)$$

independently, where $m_\alpha = m_\gamma = 0$ and $s_\alpha = s_\gamma = 1$. The t fixed effects are generated by $\lambda_t = 0$ for all t but for one year t when a universal shock of $\lambda_t = 2$ is applied. The it and jt fixed effects are

True Model = (I)		Fixed Effect Estimators			**POST**	POST
$N = 10$ ($NMT = 450$)	OLS	FE-I	FE-II	FE-III	**(2.1)**	(2.2)
Under-Fitting or Over-Fitting	Under	Correct	Over	Over	**Robust**	Robust
Average	1.317	0.997	0.997	0.999	**1.056**	1.100
Bias	0.317	-0.003	-0.003	-0.001	**0.056**	0.100
Standard Deviation	0.344	0.481	0.483	0.530	**0.401**	0.411
Root Mean Square Error	0.468	0.481	0.483	0.530	**0.405**	0.423
95% Coverage	0.836	0.942	0.941	0.909	**0.962**	0.959

True Model = (II)		Fixed Effect Estimators			**POST**	POST
$N = 10$ ($NMT = 450$)	OLS	FE-I	FE-II	FE-III	**(2.1)**	(2.2)
Under-Fitting or Over-Fitting	Under	Under	Correct	Over	**Robust**	Robust
Average	1.316	1.216	0.997	1.007	**1.117**	1.131
Bias	0.316	0.216	-0.003	0.007	**0.117**	0.131
Standard Deviation	0.351	0.463	0.483	0.529	**0.407**	0.415
Root Mean Square Error	0.472	0.511	0.483	0.529	**0.423**	0.435
95% Coverage	0.840	0.915	0.941	0.912	**0.958**	0.956

True Model = (III)		Fixed Effect Estimators			**POST**	POST
$N = 10$ ($NMT = 450$)	OLS	FE-I	FE-II	FE-III	**(2.1)**	(2.2)
Under-Fitting or Over-Fitting	Under	Under	Under	Correct	**Robust**	Robust
Average	1.619	1.603	1.551	1.053	**1.206**	1.157
Bias	0.619	0.603	0.551	0.053	**0.206**	0.157
Standard Deviation	0.347	0.355	0.370	0.518	**0.431**	0.473
Root Mean Square Error	0.710	0.700	0.663	0.521	**0.478**	0.498
95% Coverage	0.559	0.595	0.664	0.911	**0.939**	0.941

Table D.1: Monte Carlo simulation results under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with size $N = 10$ ($NMT = 450$).

True Model = (I)		Fixed Effect Estimators			**POST**	POST
$N = 15$ ($NMT = 1050$)	OLS	FE-I	FE-II	FE-III	**(2.1)**	(2.2)
Under-Fitting or Over-Fitting	Under	Correct	Over	Over	**Robust**	Robust
Average	1.283	0.995	0.995	0.995	**1.003**	1.051
Bias	0.283	-0.005	-0.005	-0.005	**0.003**	0.051
Standard Deviation	0.224	0.312	0.313	0.331	**0.270**	0.273
Root Mean Square Error	0.360	0.312	0.313	0.331	**0.270**	0.278
95% Coverage	0.745	0.943	0.942	0.928	**0.960**	0.957

True Model = (II)		Fixed Effect Estimators			**POST**	POST
$N = 15$ ($NMT = 1050$)	OLS	FE-I	FE-II	FE-III	**(2.1)**	(2.2)
Under-Fitting or Over-Fitting	Under	Under	Correct	Over	**Robust**	Robust
Average	1.281	1.233	0.995	1.000	**1.063**	1.080
Bias	0.281	0.233	-0.005	0.000	**0.063**	0.080
Standard Deviation	0.229	0.298	0.313	0.331	**0.275**	0.276
Root Mean Square Error	0.363	0.378	0.313	0.331	**0.282**	0.287
95% Coverage	0.757	0.870	0.942	0.928	**0.954**	0.952

True Model = (III)		Fixed Effect Estimators			**POST**	POST
$N = 15$ ($NMT = 1050$)	OLS	FE-I	FE-II	FE-III	**(2.1)**	(2.2)
Under-Fitting or Over-Fitting	Under	Under	Under	Correct	**Robust**	Robust
Average	1.554	1.501	1.428	0.995	**1.080**	1.085
Bias	0.554	0.501	0.428	-0.005	**0.080**	0.085
Standard Deviation	0.228	0.239	0.252	0.331	**0.284**	0.277
Root Mean Square Error	0.599	0.555	0.496	0.331	**0.295**	0.290
95% Coverage	0.324	0.448	0.598	0.928	**0.942**	0.943

Table D.2: Monte Carlo simulation results under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with size $N = 15$ ($NMT = 1050$).

generated by

$$\alpha_{it} \sim N\left(m_\alpha, s_\alpha^2 \Big/ \left(\sqrt{i} \cdot (\log(i+1))^3\right)\right)$$

and

$$\gamma_{jt} \sim N\left(m_\gamma, s_\gamma^2 \Big/ \left(\sqrt{j} \cdot (\log(j+1))^3\right)\right),$$

independently, where $m_\alpha = m_\gamma = 0$ and $s_\alpha = s_\gamma = 1$.

Table D.3 displays Monte Carlo simulation results under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with the sample size $N = 200$ ($NMT = 1900$). Observe that this table exhibits the same qualitative pattern as that in Table 1 presented in the main text, and hence we can draw the same conclusion here as well. Namely, POST (2.1) delivers a robust and superior performance compared with any other method even under the stochastic fixed effect design.

D.2.2 Alternative FE Design (2)

The i and j fixed effects are generated by $\alpha_i = (-1)^i$ for each $i \in \mathbb{N}$ and $\gamma_j = (-1)^j$ for each $j \in \mathbb{N}$. The t fixed effects are generated by $\lambda_t = 0$ for all t but for one year t when a universal shock of $\lambda_t = 2$ is applied. Note that this fixed effect design violates the approximate sparsity condition, and therefore, our proposed method is no longer guaranteed to work in this design.

Table D.4 displays Monte Carlo simulation results under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with the sample size $N = 200$ ($NMT = 1900$). Interestingly, POST (2.1) still performs at least as well as all the other estimators including the oracle one in terms of root mean square error. On the other hand, the coverage accuracy by POST (2.1) is not as well as in the baseline FE design presented in the main text.

D.2.3 Alternative FE Design (3)

The i and j fixed effects are generated by $\alpha_i \sim N\left(m_\alpha, s_\alpha^2\right)$ and $\gamma_j \sim N\left(m_\gamma, s_\gamma^2\right)$ independently, where $m_\alpha = m_\gamma = 0$ and $s_\alpha = s_\gamma = 1$. The t fixed effects are generated by $\lambda_t = 0$ for all t but for one year t when a universal shock of $\lambda_t = 2$ is applied. The it and jt fixed effects are generated by $\alpha_{it} \sim N\left(m_\alpha, s_\alpha^2\right)$ and $\gamma_{jt} \sim N\left(m_\gamma, s_\gamma^2\right)$, independently, where $m_\alpha = m_\gamma = 0$ and $s_\alpha = s_\gamma = 1$.

True Model = (I), Alternative FE (1)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Correct	Over	Over	**Robust**
Average	1.467	1.000	1.000	0.998	**0.952**
Bias	0.467	0.000	0.000	-0.002	**-0.048**
Standard Deviation	0.165	0.231	0.231	0.241	**0.198**
Root Mean Square Error	0.495	0.231	0.231	0.241	**0.204**
95% Coverage	0.203	0.947	0.946	0.935	**0.954**

True Model = (II), Alternative FE (1)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Under	Correct	Over	**Robust**
Average	1.467	1.161	1.000	1.001	**1.016**
Bias	0.467	0.161	0.000	0.001	**0.016**
Standard Deviation	0.167	0.227	0.231	0.241	**0.203**
Root Mean Square Error	0.496	0.278	0.231	0.241	**0.203**
95% Coverage	0.211	0.888	0.946	0.935	**0.958**

True Model = (III), Alternative FE (1)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Under	Under	Correct	**Robust**
Average	1.415	1.411	1.362	0.999	**1.037**
Bias	0.415	0.411	0.362	-0.001	**0.037**
Standard Deviation	0.170	0.172	0.181	0.244	**0.216**
Root Mean Square Error	0.448	0.445	0.405	0.244	**0.219**
95% Coverage	0.311	0.328	0.472	0.933	**0.946**

Table D.3: Monte Carlo simulation results for the alternative fixed effect design (1) under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with size $N = 20$ ($NMT = 1900$).

True Model = (I), Alternative FE (2)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Correct	Over	Over	**Robust**
Average	1.725	1.000	0.999	0.998	**1.060**
Bias	0.725	0.000	-0.001	-0.002	**0.060**
Standard Deviation	0.161	0.233	0.233	0.243	**0.221**
Root Mean Square Error	0.743	0.233	0.233	0.243	**0.229**
95% Coverage	0.005	0.946	0.945	0.934	**0.925**

True Model = (II), Alternative FE (2)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Under	Correct	Over	**Robust**
Average	1.725	1.112	0.999	1.000	**1.076**
Bias	0.725	0.112	-0.001	0.000	**0.076**
Standard Deviation	0.162	0.231	0.233	0.243	**0.223**
Root Mean Square Error	0.743	0.257	0.233	0.243	**0.236**
95% Coverage	0.005	0.916	0.945	0.934	**0.918**

True Model = (III), Alternative FE (2)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Under	Under	Correct	**Robust**
Average	1.563	1.482	1.314	1.002	**1.073**
Bias	0.563	0.482	0.314	0.002	**0.073**
Standard Deviation	0.175	0.186	0.204	0.243	**0.225**
Root Mean Square Error	0.589	0.516	0.374	0.243	**0.237**
95% Coverage	0.106	0.258	0.650	0.933	**0.933**

Table D.4: Monte Carlo simulation results for the alternative fixed effect design (2) under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with size $N = 20$ ($NMT = 1900$).

True Model = (I), Alternative FE (3)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Correct	Over	Over	**Robust**
Average	1.683	1.000	1.000	1.000	**1.041**
Bias	0.683	0.000	0.000	0.000	**0.041**
Standard Deviation	0.163	0.234	0.234	0.244	**0.219**
Root Mean Square Error	0.702	0.234	0.234	0.244	**0.223**
95% Coverage	0.014	0.946	0.945	0.934	**0.936**

True Model = (II), Alternative FE (3)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Under	Correct	Over	**Robust**
Average	1.683	1.117	1.000	1.002	**1.060**
Bias	0.683	0.117	0.000	0.002	**0.060**
Standard Deviation	0.163	0.232	0.234	0.244	**0.223**
Root Mean Square Error	0.702	0.259	0.234	0.244	**0.231**
95% Coverage	0.015	0.915	0.945	0.934	**0.926**

True Model = (III), Alternative FE (3)		Fixed Effect Estimators			**POST**
$N = 20$ ($NMT = 1900$)	OLS	FE-I	FE-II	FE-III	**(2.1)**
Under-Fitting or Over-Fitting	Under	Under	Under	Correct	**Robust**
Average	1.762	1.716	1.667	0.999	**1.069**
Bias	0.762	0.716	0.667	-0.001	**0.069**
Standard Deviation	0.165	0.172	0.176	0.243	**0.228**
Root Mean Square Error	0.780	0.736	0.689	0.243	**0.239**
95% Coverage	0.004	0.014	0.038	0.934	**0.929**

Table D.5: Monte Carlo simulation results for the alternative fixed effect design (3) under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with size $N = 20$ ($NMT = 1900$).

Table D.5 displays Monte Carlo simulation results under Model (I) (top panel), Model (II) (middle panel), and Model (III) (bottom panel) with the sample size $N = 200$ ($NMT = 1900$). Similarly to the case of the alternative FE design (2), POST (2.1) still performs at least as well as all the other estimators including the oracle one in terms of root mean square error. On the other hand, the coverage accuracy by POST (2.1) is not as good as in the baseline FE design presented in the main text.

References

Belloni, A., D. Chen, V. Chernozhukov and C. Hansen. 2012. "Sparse models and methods for optimal instruments with an application to eminent domain." *Econometrica* 80(6):2369–2429.

Bickel, Peter J, Ya'acov Ritov and Alexandre B Tsybakov. 2009. "Simultaneous analysis of Lasso and Dantzig selector." *The Annals of Statistics* 37(4):1705–1732.

Caner, Mehmet and Anders Bredahl Kock. 2018. "Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative Lasso." *Journal of Econometrics* 203(1):143–168.

Chernozhukov, Victor, Denis Chetverikov and Kengo Kato. 2014. "Gaussian approximation of suprema of empirical processes." *The Annals of Statistics* 42(4):1564–1597.

Chernozhukov, Victor, Denis Chetverikov and Kengo Kato. 2015. "Comparison and anti-concentration bounds for maxima of Gaussian random vectors." *Probability Theory and Related Fields* 162(1-2):47–70.

Kock, Anders Bredahl and Haihan Tang. 2019. "Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects." *Econometric Theory* 35(2):295–359.