Supplementary Material to "Nuclear Norm Regularized Quantile Regression with Interactive Fixed Effects"

Junlong Feng*

Abstract

Appendix S.A introduces an alternative set of assumptions under which the results of the paper hold without the constraint $\|L\|_\infty \leq \alpha_{NT}$ in the definition of the estimator (2.3). It also discusses the assumptions in this paper and in related literature. Appendix S.B collects all the proofs.

Appendix S.A On the Lower Bound in Theorem 1

In this appendix, we first introduce an alternative set of assumptions under which a similar quadratic lower bound as in Theorem 1 can be obtained. Under these assumptions, we can drop the constraint in the minimization problem (2.3) that defines our estimator. We then compare the assumptions in this paper with those in Ando and Bai (2020), Belloni et al. (2023) and Chen et al. (2021).

S.A.1 Dropping the Constraint in Equation (2.3)

In this section, we maintain Assumption 2 on the conditional density and add a new assumption so that the requirement $\|\hat{L}(u)\|_\infty \leq \alpha_{NT}$ can be dropped while a similar lower bound as in Theorem 1 can still be obtained. To illustrate the intuition, let us consider the case without covariates.

*Department of Economics, Hong Kong University of Science and Technology, Hong Kong SAR. E-mail address: jlfeng@ust.hk.

In the main text, we lower bound the following quantity by $\|\Delta_L\|_F^2$ multiplied by some constant for all $\Delta_L \in \mathcal{D} \coloneqq \{\Delta_L \in \mathbb{R}^{N\times T} : \|\Delta_L\|_\infty \le 2\alpha_{NT}\}$ (see equation (4.6)):

$$\sum_{i,t} \int_0^{\Delta_{L,it}} \left(F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0) \right) ds. \tag{S.A.1}$$

We only focus on $\mathcal{D}$ because we can show that the estimation error $\hat{\Delta}_L(u)$ lies in $\mathcal{D}$ uniformly in $u \in \mathcal{U}$ w.p.a.1 under the constrained parameter space $\mathcal{L}$ and by $\|L_0(u)\|_\infty \le \alpha_{NT}$.

Now that we are to replace the constrained parameter space $\mathcal{L}$ with $\mathbb{R}^{N\times T}$, $\hat{\Delta}_L(u)$ may no longer lie in $\mathcal{D}$ w.p.a.1. We need to constrain $\hat{\Delta}_L$ in a different set. For any $\Delta_L \in \mathbb{R}^{N\times T}$, let $\mathcal{P}_\Omega \Delta_L$ be an $N \times T$ matrix whose (i,t)-th element is $\mathbb{1}(|\Delta_{L,it}| \le 2\alpha_{NT}) \cdot \Delta_{L,it}$. Let $\mathcal{P}_{\Omega^\perp}\Delta_L \coloneqq \Delta - \mathcal{P}_\Omega \Delta_L$. By construction, $\|\Delta_L\|_F^2 = \|\mathcal{P}_\Omega \Delta_L\|_F^2 + \|\mathcal{P}_{\Omega^\perp}\Delta_L\|_F^2$. Let $\|\mathcal{P}_\Omega \Delta_L\|_F^2 = C_{sm}\|\Delta_L\|_F^2$ where C_{sm} is in $[0,1]$ and may depend on N and T. Note that $\mathcal{D}$ is equivalent to the set of matrices whose C_{sm} equals 1. Quantity (S.A.1) is equal to

$$\begin{aligned}
&\sum_{i,t} \int_0^{\Delta_{L,it}} \left(F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0) \right) ds \\
=& \sum_{\{i,t: |\Delta_{L,it}| \le 2\alpha_{NT}\}} \int_0^{\Delta_{L,it}} \left(F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0) \right) ds \\
&+ \sum_{\{i,t: |\Delta_{L,it}| > 2\alpha_{NT}\}} \int_0^{\Delta_{L,it}} \left(F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0) \right) ds.
\end{aligned}$$

For the first sum on the right side, we can lower bound it by $C_{min}\|\mathcal{P}_\Omega \Delta_L\|_F^2/\alpha_{NT}^2$ for some $C_{min} > 0$ (the proof is similar to that of Theorem 1 and is thus omitted). For the second term, now that $\Delta_{L,it}$ can be unbounded, the conditional density $f_{V_{it}(u)}$ may be arbitrarily close or equal to zero. Hence, it can only be lower bounded by 0. Yet as long as $\|\mathcal{P}_\Omega \Delta_L\|_F^2$ is of a nonnegligible proportion of $\|\Delta_L\|_F^2$, we can still lower bound (S.A.1) be $\|\Delta\|_F^2$ multiplied by some constant.

Formally, assume there exists a constant $C_{sm} > 0$ such that for all $u \in \mathcal{U}$, we have $\hat{\Delta}_L(u) \in \mathcal{D}^{(2)}$ w.p.a.1 where $\mathcal{D}^{(2)}$ is the following cone[1]

$$\mathcal{D}^{(2)} \coloneqq \left\{ \Delta_L \in \mathbb{R}^{N\times T} : \|\mathcal{P}_\Omega \Delta_L\|_F^2 \ge C_{sm}\|\Delta_L\|_F^2 \right\}. \tag{S.A.2}$$

[1]One can alternatively impose the restriction on the space where $\hat{\Delta}_L(u)$ lies w.p.a.1, instead of on $\hat{\Delta}_L(u)$ directly. For instance, assume that any $\Delta_L \in \mathcal{R}_u$ with $\|\Delta_L\|_F \le \sqrt{NT}\gamma$ satisfies $\Delta_L \in \mathcal{D}^{(2)}$. The proof of uniform consistency still goes through because $\hat{\Delta}_L(u) \in \mathcal{R}_u$ w.p.a.1 and in the proof we only focus on the sphere $\|\Delta_L\|_F = \sqrt{NT}\gamma$.

We can then restrict our analysis within $\mathcal{D}^{(2)}$ and lower bound (S.A.1) for all $\Delta_L \in \mathcal{D}^{(2)}$ by

$$C_{min}\|\mathcal{P}_\Omega \Delta_L\|_F^2/\alpha_{NT}^2 + 0 \geq C_{sm}C_{min}\|\Delta_L\|_F^2/\alpha_{NT}^2.$$

Then we can obtain an error bound on the estimator which has the same order as that in Theorem 2 since the quadratic lower bound has the same order.

Both $\mathcal{D}$ (adopted in the main text) and $\mathcal{D}^{(2)}$ here limit the spikiness of the matrices Δ_Ls in them. Set $\mathcal{D}$ restricts the *magnitude* of the large entries in $\Delta_L \in \mathcal{D}$. In contrast, by definition (S.A.2), set $\mathcal{D}^{(2)}$ restricts both the magnitude and the *number* of large entries in $\Delta_L \in \mathcal{D}^{(2)}$. In particular, on the sphere $\|\Delta_L\|_F^2 = NT\gamma^2$ where γ is the same as in Section 3 with $p = 0$, $\mathcal{D}^{(2)}$ restricts the number of large entries in the sense that entries in $\Delta_L \in \mathcal{D}^{(2)}$ on this sphere can be as large as $\sqrt{(1 - C_{sm})NT}\gamma$, but for any $\delta_{NT} > 2\alpha_{NT}$, the number of entries whose magnitude are equal to δ_{NT} is at most $(1 - C_{sm})NT\gamma^2/\delta_{NT}^2$.

When there are covariates, complications arise and C_{sm} not only needs to be bounded away from zero but also needs to be sufficiently large. A sufficient condition is that $C_{sm} \to 1$ as N and T grow to infinity. Specifically, we have the following theorem.

Theorem S.A.1. *Let λ be the same as in Lemma 1. Under Assumptions 1 to 5, if w.p.a.1, $\hat{\Delta}_L(u) \in \mathcal{D}^{(2)}$ defined in equation (S.A.2) with $C_{sm} \to 1$ as N and T grow to infinity, then for some constant $C_{error,2} > 0$, the following estimator*

$$(\hat{\beta}(u), \hat{L}(u)) = \arg \min_{\beta \in \mathbb{R}^p, L \in \mathbb{R}^{N \times T}} \frac{1}{NT} \boldsymbol{\rho}_u(Y - \sum_{j=1}^{p} X_j \beta_j - L) + \lambda \|L\|_*$$

satisfies the inequality below w.p.a.1.

$$\begin{aligned}
&\sup_{u \in \mathcal{U}} \|\hat{\beta}(u) - \beta_0(u)\|_F^2 + \frac{1}{NT}\|\hat{L}(u) - L_0(u)\|_F^2 \\
&\leq C_{error,2}^2 \alpha_{NT}^4 \left[(1 + C_\lambda)^2 \vee \log(NT)\right] \left(\frac{p \log((p+1)NT)}{NT} \vee \frac{\bar{r}}{N \wedge T}\right).
\end{aligned}$$

Proof. See Appendix S.B.4. ∎

Remark S.A.1. *The difference between the estimator defined in Theorem S.A.1 and the one defined by equation (2.3) lies in the parameter space of L.*

S.A.2 Comparison of Different Approaches

In this section, we compare our assumptions to obtain a quadratic lower bound in Theorem 1 with those in Ando and Bai (2020), Belloni et al. (2023) and Chen et al. (2021). To highlight the differences, we still consider the case where there are no covariates. Also, since these mentioned papers all focus on consistency pointwise in $u \in \mathcal{U}$, in the following discussion we also drop the requirements on uniformity in u in our assumptions.

First, let us summarize the assumptions needed in our two approaches to a quadratic lower bound. Recall that without the covariates, the set $\mathcal{D}$ in the main text is defined as $\{\Delta_L \in \mathbb{R}^{N\times T} : \|\Delta_L\|_\infty \leq 2\alpha_{NT}\}$.

- Approach 1 (adopted in the main text).
  - On the conditional density of $V_{it}(u)$: Assumption 2.
  - On the *magnitude* of large elements in $\hat{\Delta}_L(u)$: Assuming $\|L_0(u)\|_\infty \leq \alpha_{NT}$ in equation (3.1) and by the constraint in the estimator's definition (2.3), we have $\hat{\Delta}_L(u) \in \mathcal{D}$ for all $u \in \mathcal{U}$.
- Approach 2 (introduced in Appendix S.A.1).
  - On the conditional density of $V_{it}(u)$: Assumption 2.
  - On the *number* and *magnitude* of large elements in $\hat{\Delta}_L(u)$: Assume $\|L_0(u)\|_\infty \leq \alpha_{NT}$ and $\hat{\Delta}_L(u) \in \mathcal{D}^{(2)}$ for all $u \in \mathcal{U}$ w.p.a.1 with $C_{sm} \to 1$.

Ando and Bai (2020) and Chen et al. (2021) impose stronger assumptions on density $f_{V_{it}(u)}$. They both assume that $f_{V_{it}(u)}(\cdot)$ is continuous and for *any* compact set S, there exists an S-dependent constant $\underline{f}_S$ such that the density $f_{V_{it}(u)}(s) \geq \underline{f}_S > 0$ for all $s \in S$ and all i and t. Note that this assumption implies our Assumption 2 by choosing $S = [-\delta, \delta]$ for any $\delta > 0$. This stronger assumption can help obtain a quadratic lower bound for our purpose by a simpler argument if $\|L_0(u)\| \leq \alpha_{NT}$ and $\|L\|_\infty \leq \alpha_{NT}$ are still imposed[2]. To see this, by $|\Delta_{L,it}| \leq 2\alpha_{NT}$, their assumption implies that there exists a constant $\underline{f}_{\alpha_{NT}} > 0$ such that (S.A.1) is lower bounded by $\underline{f}_{\alpha_{NT}} \sum_{i,t} \Delta^2_{L,it}/2$ by directly applying first-order Taylor expansion. Similar to our Approach 1, the lower bound also depends on α_{NT} via $\underline{f}_{\alpha_{NT}}$.

Now let us turn to Belloni et al. (2023). Their approach is more similar to our Approach 2 because they also restrict the magnitude and the number of large entries in $\hat{\Delta}_L(u)$. Again, since they only focus on pointwise consistency, we compare our related assumptions with theirs by dropping the required uniformity in u. Like our approaches, their assumptions also consist of two parts:

First, on the conditional density of $V_{it}(u)$, their Assumption 3 requires that for all i and t, the density $f_{V_{it}(u)}(v)$ is bounded away from 0 at $v = 0$ by $\underline{f}$ and bounded from above uniformly in v. Meanwhile, $f_{V_{it}(u)}$ is assumed to be differentiable and its derivative

[2] Indeed, these two papers assume elements in $L_0(u)$ lie in a fixed compact space, i.e. α_{NT} is fixed, not (N, T)-dependent.

$\partial f_{V_{it}(u)}(v)/\partial v$ is continuous and bounded in absolute value by $\bar{f}'$ uniformly in v, i, and t. These two requirements are stronger than our Assumption 2 since uniform boundedness of both a set of functions and of their derivatives implies equicontinuity.

Second, on the magnitude and the number of large entries in $\hat{\Delta}_L(u)$, their Assumption 4 and equation (19) essentially say that $\hat{\Delta}_L(u) \in \mathcal{D}^{(3)}$ w.p.a.1 where[3]

$$\mathcal{D}^{(3)} \in \left\{ \Delta_L \in \mathbb{R}^{N\times T} : \frac{\underline{f}}{2}\|\Delta_L\|_F^2 - \frac{\bar{f}'}{3}\sum_{i,t}|\Delta_{L,it}|^3 \geq 0 \right\} \tag{S.A.3}$$

and the constants $\underline{f}$ and $\bar{f}'$ are introduced in the previous paragraph. By the inequality in (S.A.3) and by their assumption on the conditional density, they lower bound (S.A.1) for $\Delta_L \in \mathcal{D}^{(3)}$ by second-order Taylor expansion:

$$\sum_{i,t}\int_0^{\Delta_{L,it}} \left(F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0)\right) ds \geq \frac{\underline{f}}{4}\|\Delta\|_F^2 + \left(\frac{\underline{f}}{4}\|\Delta\|_F^2 - \frac{\bar{f}'}{6}\sum_{i,t}|\Delta_{L,it}|^3\right) \geq \frac{\underline{f}}{4}\|\Delta\|_F^2.$$

The set $\mathcal{D}^{(3)}$ serves a similar purpose as $\mathcal{D}^{(2)}$ in our Approach 2. Both sets restrict the magnitude and the number of large entries in the matrices in these sets. Yet the condition on $\mathcal{D}^{(3)}$ can be more restrictive in the sense that large elements allowed in $\mathcal{D}^{(3)}$ can be fewer than $\mathcal{D}^{(2)}$. To see this, suppose $\|\Delta_L\|_F$ has order ν_{NT} and let $\alpha_{NT} \to \infty$ and $\alpha_{NT} = o(\nu_{NT})$. For large element $\Delta_{L,it}$ of order $\delta_{NT} > 2\alpha_{NT}$, in any matrix in $\mathcal{D}^{(2)}$, there can be as many as $o(\nu_{NT}^2/\delta_{NT}^2)$ of such elements while $C_{sm} \to 1$ still holds. But in any matrix in $\mathcal{D}^{(3)}$, there can be only $O(\nu_{NT}^2/\delta_{NT}^3)$ of them.

Comparing $\mathcal{D}$, $\mathcal{D}^{(2)}$ and $\mathcal{D}^{(3)}$, note that $\hat{\Delta}_L(u) \in \mathcal{D}$ for all $u \in \mathcal{U}$ w.p.a.1 can be guaranteed under primitive conditions: $\|L_0(u)\|_\infty \leq \alpha_{NT}$ and the constraint in the definition of the estimator (2.3). However, $\hat{\Delta}_L(u) \in \mathcal{D}^{(2)}$ in our Approach 2 and $\hat{\Delta}_L(u) \in \mathcal{D}^{(3)}$ in Belloni et al. (2023) are higher level conditions.

To sum up, to obtain a quadratic lower bound in Theorem 1, we need to i) make assumptions on the conditional density $f_{V_{it}(u)}$ and ii) to restrict the *magnitude* of large elements in $\hat{\Delta}_L(u)$ (Approach 1) or both the *number* and the magnitude of them (Approach 2). Our assumption on the conditional density seems to be the weakest in the discussed literature. For large elements in $\hat{\Delta}_L(u)$, our Approach 1 is under more primitive conditions while the restriction $\mathcal{D}^{(2)}$ in our Approach 2 can be milder than the restriction $\mathcal{D}^{(3)}$ in Belloni et al. (2023). On the other hand, in both of our two approaches, we need to assume $\|L_0(u)\|_\infty \leq \alpha_{NT}$,

[3]More specifically, in a similar way to footnote 1, the assumption is imposed on all $\Delta_L \in \mathcal{R}_u$; it basically requires $\gamma/4 \leq 3\underline{f}\|\Delta_L\|_F^3/(8\sqrt{NT}\bar{f}'\sum_{i,t}|\Delta_{L,it}|^3)$ (supposing there are no covariates). Under this assumption, $\Delta_L \in \mathcal{D}^{(3)}$ as long as $\Delta_L \in \mathcal{R}_u$ with $\|\Delta_L\|_F \leq \sqrt{NT}\gamma$.

while the approach in Belloni et al. (2023) is free of it. Finally, since they focus on high-dimensional regressors, some discussed relaxations in our approaches may not apply there. We view all three approaches as complementary.

Appendix S.B Proofs

S.B.1 Proofs of the Results in Section 3

Proof of Lemma 1

Recall that we suppress conditioning on W_L for simplicity and $V(u) \coloneqq Y - q_{Y|W_X}(u)$. Let $\nabla \boldsymbol{\rho}_u(V(u))$ be an $N \times T$ subgradient matrix of $\boldsymbol{\rho}_u(\cdot)$ evaluated at $V(u)$. Note that for check function $\rho_u(\cdot)$, $\nabla \rho_u(0)$ is not unique due to nonsmoothness of ρ_u at 0 and can be any number that lies in $[u-1, u]$. We set $\nabla \rho_u(0) = u$ for convenience. With probability one, the (i,t)-th element of $\nabla \boldsymbol{\rho}_u(V(u))$ is

$$(\nabla \boldsymbol{\rho}_u (V(u)))_{it} = u\mathbb{1}(V_{it}(u) \geq 0) + (u-1)\mathbb{1}(V_{it}(u) < 0).$$

These elements are bounded and independent with mean 0 conditional on W_X (and implicitly conditional on the fixed effects) by Assumption 1 and by the definition of $V(u)$ [4]. We introduce the following lemma for $\nabla \boldsymbol{\rho}_u(V(u))$. The proof is in Appendix S.B.4.

Lemma S.B.1. *Under Assumption 1, there exists a constant $C_{op} > 6$ such that the following inequalities hold w.p.a.1:*

$$\sup_{u\in\mathcal{U}} \max_{1\leq j\leq p} \left|\left\langle \nabla \boldsymbol{\rho}_u(V(u)), X_j\right\rangle\right| \leq 5\sqrt{2C_X NT \log((p+1)NT)}, \tag{S.B.1}$$

$$\sup_{u\in\mathcal{U}} \|\nabla \boldsymbol{\rho}_u(V(u))\| \leq C_{op}\sqrt{N \vee T}, \tag{S.B.2}$$

where C_X is defined in Assumption 1.

In what follows, the derivation is under the event that inequalities (S.B.1) and (S.B.2) hold. Since $\|L_0(u)\|_\infty \leq \alpha_{NT}$ for all $u \in \mathcal{U}$, $L_0(u)$ is a feasible solution to the minimization problem (2.3). Then by the definition of $(\hat{\beta}(u), \hat{L}(u))$, the following inequality holds with probability one:

$$\sup_{u\in\mathcal{U}} \left(\frac{1}{NT} \left[\boldsymbol{\rho}_u \left(V(u) - \sum_{j=1}^{p} X_j \hat{\Delta}_{\beta,j}(u) - \hat{\Delta}_L(u) \right) - \boldsymbol{\rho}_u(V(u)) \right] + \lambda \left(\|\hat{L}(u)\|_* - \|L_0(u)\|_* \right) \right) \leq 0, \tag{S.B.3}$$

[4] Conditional mean zero is obtained by noting that $\Pr(V_{it}(u) < 0 | W_X) = u$ almost surely by definition.

where $\hat{\Delta}_{\beta,j}(u) := \hat{\beta}_j(u) - \beta_{0,j}(u)$ and $\hat{\Delta}_L(u) := \hat{L}(u) - L_0(u)$.

Let us first consider $\left[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j\hat{\Delta}_{\beta,j}(u) - \hat{\Delta}_L(u)\Big) - \boldsymbol{\rho}_u(V(u))\right]/NT$. With probability one,

$$\begin{aligned}
&\frac{1}{NT}\left[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j\hat{\Delta}_{\beta,j}(u) - \hat{\Delta}_L(u)\Big) - \boldsymbol{\rho}_u(V(u))\right] \\
&\geq -\frac{1}{NT}\Big|\Big\langle\nabla\boldsymbol{\rho}_u(V(u)), \sum_{j=1}^p X_j\hat{\Delta}_{\beta,j}(u) + \hat{\Delta}_L(u)\Big\rangle\Big| \\
&\geq -\frac{1}{NT}\|\hat{\Delta}_\beta(u)\|_1 \max_{1\leq j\leq p}\Big|\Big\langle\nabla\boldsymbol{\rho}_u(V(u)), X_j\Big\rangle\Big| - \frac{1}{NT}\|\nabla\boldsymbol{\rho}_u(V(u))\|\cdot\|\hat{\Delta}_L(u)\|_* \\
&\geq -5\sqrt{\frac{2C_X\log((p+1)NT)}{NT}}\|\hat{\Delta}_\beta(u)\|_1 - \frac{C_{op}\sqrt{N\vee T}}{NT}\|\hat{\Delta}_L(u)\|_* \\
&\geq -5\sqrt{\frac{2C_Xp\log((p+1)NT)}{NT}}\|\hat{\Delta}_\beta(u)\|_F - \frac{C_{op}\sqrt{N\vee T}}{NT}\|\hat{\Delta}_L(u)\|_*. \qquad \text{(S.B.4)}
\end{aligned}$$

The first inequality is by the definition of subgradient. The first term in the second inequality is elementary. The second term is from Lemma 3.2 in Candès and Recht (2009) which says for any two matrices A and B of the same size, $|\langle A, B\rangle| \leq \|A\|\cdot\|B\|_*$. The penultimate inequality is by inequalities (S.B.1) and (S.B.2) in Lemma S.B.1.

Next, consider $\lambda\Big(\|\hat{L}(u)\|_* - \|L_0(u)\|_*\Big)$. Recall that $\mathcal{P}_{\Phi(u)^\perp}$ is the orthogonal projection onto the orthogonal complement of $\Phi(u)$. By construction, $\mathcal{P}_{\Phi(u)^\perp}L_0(u) = 0$. Moreover, for any $N\times T$ matrix M, $\|\mathcal{P}_{\Phi(u)}M + \mathcal{P}_{\Phi(u)^\perp}M\|_* = \|\mathcal{P}_{\Phi(u)}M\|_* + \|\mathcal{P}_{\Phi(u)^\perp}M\|_*$ since $\mathcal{P}_{\Phi(u)}M$ and $\mathcal{P}_{\Phi(u)^\perp}M$ have orthogonal singular vectors to each other. Hence, by $\hat{L}(u) = L_0(u) + \hat{\Delta}_L(u)$, with probability one,

$$\begin{aligned}
\|\hat{L}(u)\|_* - \|L_0(u)\|_* =& \|\mathcal{P}_{\Phi(u)}L_0(u) + \mathcal{P}_{\Phi(u)}\hat{\Delta}_L(u)\|_* + \|\mathcal{P}_{\Phi(u)^\perp}\hat{\Delta}_L(u)\|_* - \|\mathcal{P}_{\Phi(u)}L_0(u)\|_* \\
\geq& \|\mathcal{P}_{\Phi(u)^\perp}\hat{\Delta}_L(u)\|_* - \|\mathcal{P}_{\Phi(u)}\hat{\Delta}_L(u)\|_*. \qquad \text{(S.B.5)}
\end{aligned}$$

Combining equations (S.B.3), (S.B.4) and (S.B.5), we have shown that

$$\begin{aligned}
\sup_{u\in\mathcal{U}}\Bigg(&\left(\lambda - \frac{C_{op}\sqrt{N\vee T}}{NT}\right)\|\mathcal{P}_{\Phi(u)^\perp}\hat{\Delta}_L(u)\|_* - 5\sqrt{\frac{2C_Xp\log((1+p)NT)}{NT}}\|\hat{\Delta}_\beta(u)\|_F \\
&- \left(\lambda + \frac{C_{op}\sqrt{N\vee T}}{NT}\right)\|\mathcal{P}_{\Phi(u)}\hat{\Delta}_L(u)\|_*\Bigg) \leq 0.
\end{aligned}$$

So, we have

$$\sup_{u\in\mathcal{U}}\Bigg(\Big\|\mathcal{P}_{\Phi(u)^\perp}\hat{\Delta}_L(u)\Big\|_* - \frac{NT\lambda + C_{op}\sqrt{N\vee T}}{NT\lambda - C_{op}\sqrt{N\vee T}}\Big\|\mathcal{P}_{\Phi(u)}\hat{\Delta}_L(u)\Big\|_*$$

$$-\frac{5\sqrt{2C_X pNT\log((p+1)NT)}}{NT\lambda - C_{op}\sqrt{N\vee T}}\left\|\hat{\Delta}_{\beta}(u)\right\|_F\Bigg) \leq 0$$

holding with probability one under the event that equations (S.B.1) and (S.B.2) hold, which in turn hold w.p.a.1 by Lemma S.B.1. We obtain the desired results by substituting $\lambda = (1+C_\lambda)C_{op}\sqrt{N\vee T}/NT$ into it. ■

S.B.2 Proof of the Results in Section 4

Proof of Equivalence between Assumption 4 and Equation (4.4)

Noting that both $M_{\Lambda_0}(u)$ and $M_{F_0}(u)$ are idempotent, we have

$$\begin{aligned}
\mathbb{E}\left\|M_{\Lambda_0}(u)\left(\sum_{j=1}^{p}X_j\tau_j\right)M_{F_0}(u)\right\|_F^2 =& \mathbb{E}\mathrm{Tr}\left[M_{\Lambda_0}(u)\left(\sum_{j=1}^{p}X_j\tau_j\right)M_{F_0}(u)M_{F_0(u)}\left(\sum_{j=1}^{p}X_j\tau_j\right)'M_{\Lambda_0}(u)\right]\\
=& \mathbb{E}\mathrm{Tr}\left[M_{\Lambda_0}(u)\left(\sum_{j=1}^{p}X_j\tau_j\right)M_{F_0}(u)\left(\sum_{j=1}^{p}X_j\tau_j\right)'\right]\\
=& \tau'\mathbb{E}\left(x'\left(M_{F_0}(u)\otimes M_{\Lambda_0}(u)\right)x\right)\tau,
\end{aligned}$$

where Tr is the trace of a matrix. The first equality is by definition of the Frobenius norm. The last equality follows equation (S.2) in Moon and Weidner (2015). Then we have

$$\begin{aligned}
\mathbb{E}\left\|\mathcal{P}_{\Phi(u)^\perp}\left(\sum_{j=1}^{p}X_j\tau_j\right)\right\|_F^2 =& \mathbb{E}\left\|M_{\Lambda_0}(u)\left(\sum_{j=1}^{p}X_j\tau_j\right)M_{F_0}(u)\right\|_F^2\\
=& \tau'\mathbb{E}\left(x'\left(M_{F_0}(u)\otimes M_{\Lambda_0}(u)\right)x\right)\tau,
\end{aligned} \tag{S.B.6}$$

where the first equality is by Remark 6. Similarly,

$$\begin{aligned}
\mathbb{E}\left\|\mathcal{P}_{\Phi(u)}\left(\sum_{j=1}^{p}X_j\tau_j\right)\right\|_F^2 =& \mathbb{E}\left\|\sum_{j=1}^{p}X_j\tau_j\right\|_F^2 - \mathbb{E}\left\|\mathcal{P}_{\Phi(u)^\perp}\left(\sum_{j=1}^{p}X_j\tau_j\right)\right\|_F^2\\
=& \tau'\mathbb{E}(x'x)\tau - \tau'\mathbb{E}\left(x'\left(M_{F_0}(u)\otimes M_{\Lambda_0}(u)\right)x\right)\tau.
\end{aligned} \tag{S.B.7}$$

Substituting equations (S.B.6) and (S.B.7) into the left side of equation (4.2) in Assumption 4, equation (4.2) then holds if and only if equation (4.4) holds.

Proof Sketch of Theorem 1

The proof of Theorem 1 (presented in the next section) is long and involves quite a bit of algebraic detail. For the convenience of the reader, we first sketch the proof to highlight the key steps.

We first prove a lemma which lower bounds the expectation in the theorem by a linear function of $\mathbb{E}\left\|\sum_{j=1}^{p} X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2$ (see Lemma S.B.2 for details). Then it is sufficient to show that $\mathbb{E}\left\|\sum_{j=1}^{p} X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2 \geq C_{RSC}\left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right)$ for all $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$ for some $C_{RSC} > 0$ and Theorem 1 follows. We distinguish two cases. The first case is $r(u) = 0$. In this case, for any $N \times T$ matrix M, $\mathcal{P}_{\Phi(u)}M = 0$. Therefore, if $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$, it is necessary that $\|\Delta_L\|_F^2 = \|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\|_F^2 \leq \kappa_2(\lambda)^2\|\Delta_\beta\|_F^2$ by the definition of $\mathcal{R}_u$. Assumption 3 implies that the number of regressors $p = o((N \wedge T)/\log(NT)\alpha_{NT}^2)$ and thus $\kappa_2(\lambda) = o(N \wedge T)$ by the definition of $\kappa_2(\lambda)$ in equation (3.7). So, $\|\Delta_L\|_F^2$ is $o(N^2 \wedge T^2)\|\Delta_\beta\|_F^2$. On the other hand, Assumption 4 ii) implies that $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}\|_F^2$ is no smaller than $\sigma_{min}^2 NT\|\Delta_\beta\|_F^2$. Therefore, $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}\|_F^2$ is greater than $\|\Delta_L\|_F^2$ in order. So, the quantity $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j} + \Delta_L\|_F^2$ must be bounded from below by a large fraction of $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}\|_F^2$, and consequently by (still a large fraction of) $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}\|_F^2 + \|\Delta_L\|_F^2$ because adding $\|\Delta_L\|_F^2$ does not change the order.

Then we move to the more interesting case of $r(u) \geq 1$. Recall that the definition of $\mathcal{R}_u$ implies that $\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\|_F \leq \sqrt{3r(u)}\kappa_1(\lambda)|\mathcal{P}_{\Phi(u)}\Delta_L\|_F + \kappa_2(\lambda)\|\Delta_\beta\|_F$. Similar to the $r(u) = 0$ case, a simple case is when $\sqrt{3r(u)}\kappa_1(\lambda)\|\mathcal{P}_{\Phi(u)}\Delta_L\|_F \leq \kappa_2(\lambda)\|\Delta_\beta\|_F$. In this case, the order of $\|\Delta_L\|_F$ is bounded by $\kappa_2(\lambda)\|\Delta_\beta\|_F$ because both the order of $\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\|_F$ and of $\|\mathcal{P}_{\Phi(u)}\Delta_L\|_F$ are bounded by it. Again, this is dominated by the order of $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}\|_F^2$ following the argument as the $r(u) = 0$ case.

The nontrivial case is when $r(u) \geq 1$ and $\sqrt{3r(u)}\kappa_1(\lambda)\|\mathcal{P}_{\Phi(u)}\Delta_L\|_F > \kappa_2(\lambda)\|\Delta_\beta\|_F$. In this case, it is possible that $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}\|_F^2$ and $\|\Delta_L\|_F^2$ are of the same order, and consequently, without further restriction, the quantity of interest $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j} + \Delta_L\|_F^2$ can reach 0 even when Δ_L and Δ_β are not. We then invoke Assumption 4 i), the key identification condition not used yet, to separate the two matrices out. We show that this assumption guarantees the existence of two positive constants $c_1(u), c_2(u)$ with $c_1(u)c_2(u) \leq 1/C_{\Phi X} < 1$ such that $\mathbb{E}\|\sum_{j=1}^{p} X_j\Delta_{\beta,j} + \Delta_L\|_F^2$ is bounded from below by

$$\left(c_1(u)\sqrt{\mathbb{E}\left\|\mathcal{P}_{\Phi(u)^\perp}\left(\sum_{j=1}^{p} X_j\Delta_{\beta,j}\right)\right\|_F^2} - \left(\kappa_2(\lambda)\frac{\|\Delta_\beta\|_F}{\sqrt{3r(u)}\kappa_1(\lambda)} + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)\right)^2$$

$$+\left(\sqrt{\mathbb{E}\left\|\mathcal{P}_{\Phi(u)^\perp}\left(\sum_{j=1}^p X_j\Delta_{\beta,j}\right)\right\|_F^2}-c_2(u)\left(\kappa_2(\lambda)\frac{\|\Delta_\beta\|_F}{\sqrt{3r(u)\kappa_1(\lambda)}}+\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)\right)^2.$$

Observing this quantity, we notice that by $c_1(u)c_2(u)$ being bounded away from 1, the two quadratic terms *cannot* be simultaneously equal to 0 unless every term in them is 0. This is the source of identification. We then prove that it is further bounded from below by a positive fraction of $\mathbb{E}\|\sum_{j=1}^p X_j\Delta_{\beta,j}\|_F^2+\|\Delta_L\|_F^2$.

Proof of Theorem 1

Recall that $\mathcal{D}:=\mathbb{R}^p\times\{\Delta_L\in\mathbb{R}^{N\times T}:\|\Delta_L\|_\infty\leq 2\alpha_{NT}\}$. We have the following lemma.

Lemma S.B.2. *Let* $\varepsilon_{NT}:=2\left(\max_{i,t}\mathbb{E}\left(\mathbb{1}\left(\sqrt{X_{it}'X_{it}}>\alpha_{NT}/\gamma\right)\cdot X_{it}'X_{it}\right)\vee\max_{i,t}\mathbb{P}\left(\sqrt{X_{it}'X_{it}}>\alpha_{NT}/\gamma\right)\right)$. *Under Assumption 2, if* $\alpha_{NT}\geq 1$*, then the following inequality holds for any fixed* N *and* T*:*

$$\inf_{\substack{u\in\mathcal{U}\\(\Delta_\beta,\Delta_L)\in\mathcal{D}}}\left(\mathbb{E}\left(\boldsymbol{\rho}_u\Big(V(u)-\sum_{j=1}^p X_j\Delta_{\beta,j}-\Delta_L\Big)-\boldsymbol{\rho}_u(V(u))\right)\right.$$
$$\left.-\frac{(1\wedge\delta)^2\underline{f}\left(\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j}+\Delta_L\right\|_F^2-\varepsilon_{NT}\left(NT\|\Delta_\beta\|_F^2+\|\Delta\|_F^2\right)\right)}{2(2\alpha_{NT}+(\alpha_{NT}\|\Delta_\beta\|_F/\gamma\vee 1))^2}\right)\geq 0.\quad\text{(S.B.8)}$$

Sequence ε_{NT} *is equal to 0 for sufficiently large* N *and* T *under Assumption 3 i), and converges to zero as* $N,T\to\infty$ *under Assumption 3 ii).*

Proof. See Appendix S.B.4. ∎

Remark S.B.1. *When deriving the uniform rate of convergence later, we only need to focus on the sphere* $NT\|\Delta_\beta\|_F^2+\|\Delta_L\|_F^2=NT\gamma^2$*. Then equation (S.B.8) can be simplified as:*

$$\inf_{\substack{u\in\mathcal{U}\\\|\Delta_\beta\|_F\leq\gamma\\(\Delta_\beta,\Delta_L)\in\mathcal{D}}}\left(\mathbb{E}\left(\boldsymbol{\rho}_u\Big(V(u)-\sum_{j=1}^p X_j\Delta_{\beta,j}-\Delta_L\Big)-\boldsymbol{\rho}_u(V(u))\right)\right.$$
$$\left.-\frac{C_{min}}{\alpha_{NT}^2}\left(\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j}+\Delta_L\right\|_F^2-\varepsilon_{NT}\left(NT\|\Delta_\beta\|_F^2+\|\Delta_L\|_F^2\right)\right)\right)\geq 0,\quad\text{(S.B.9)}$$

where $C_{min}:=(1\wedge\delta)^2\underline{f}/18$*,* $\underline{f}$ *and* δ *are defined in Assumption 2, and* γ *is defined in equation (3.4).*

Now we derive a lower bound on $\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j}+\Delta_L\right\|_F^2$. For the ease of notation, in the proof we denote $\kappa_1(\lambda)$ and $\kappa_2(\lambda)$, defined by equation (3.7) in Lemma 1, by κ_1 and

κ_2. We have $\kappa_1 = O(1)$ and $\kappa_1 > 1$. Under Assumption 3, $\kappa_2 = o(N \wedge T)$ uniformly in $u \in \mathcal{U}$. When $r(u) \geq 1$, let $\kappa_3 := \frac{\kappa_2}{\sqrt{3r(u)\kappa_1}}$. So κ_3 has the same order as κ_2 uniformly in $u \in \{u \in \mathcal{U} : r(u) \geq 1\}$. We discuss the case of $r(u) = 0$ and the case of $r(u) \geq 1$ separately.

Case 1. $r(u) = 0$. When $r(u) = 0$, for any $N \times T$ matrix M, $\mathcal{P}_{\Phi(u)} M = 0$ and $\mathcal{P}_{\Phi(u)^\perp} M = M$. Hence, for $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$, $\|\Delta_L\|_F \leq \kappa_2 \|\Delta_\beta\|_F$. Meanwhile, by Assumption 4 ii), we have the following,

$$\mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\right\|_F^2 = \Delta_\beta' \left(\sum_{i,t} \mathbb{E}\left(X_{it} X_{it}'\right)\right) \Delta_\beta \geq NT\sigma_{min}^2 \|\Delta_\beta\|_F^2. \tag{S.B.10}$$

Therefore, by $\kappa_2 = o(N \wedge T)$, for an arbitrarily small $\varepsilon_1 > 0$, there exist N_1 and T_1 such that for all $N > N_1$ and $T > T_1$:

$$\|\Delta\|_F \leq \kappa_2 \|\Delta_\beta\|_F \leq \frac{\varepsilon_1}{2} \sqrt{NT\sigma_{min}^2} \|\Delta_\beta\|_F \leq \frac{\varepsilon_1}{2} \sqrt{\mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\right\|_F^2}. \tag{S.B.11}$$

Since the order of κ_2 is the same for all u, N_1 and T_1 are the same for all u as well. Therefore, the following holds for all $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$ and all $u \in \{u \in \mathcal{U} : r(u) = 0\}$,

$$\begin{aligned}
\mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L\right\|_F^2 \geq & \mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\right\|_F^2 + \|\Delta_L\|_F^2 - 2\sqrt{\mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\right\|_F^2} \|\Delta\|_F \\
\geq & (1 - \varepsilon_1) \mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\right\|_F^2 + \|\Delta_L\|_F^2 \\
\geq & \left[(1 - \varepsilon_1)\sigma_{min}^2 \wedge 1\right] \left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right),
\end{aligned}$$

where the first inequality is by the following fact: For any two (random) matrices D_1 and D_2 of the same dimensions, the following holds (with probability 1) by the Cauchy-Schwarz inequality:

$$\mathbb{E}\|D_1 + D_2\|_F^2 = \mathbb{E}\|D_1\|_F^2 + \mathbb{E}\|D_2\|_F^2 + 2\mathbb{E}\langle D_1, D_2\rangle \geq \mathbb{E}\|D_1\|_F^2 + \mathbb{E}\|D_2\|_F^2 - 2\sqrt{\mathbb{E}\|D_1\|_F^2 \mathbb{E}\|D_2\|_F^2}. \tag{S.B.12}$$

Hence, for all $N > N_1$ and $T > T_1$,

$$\inf_{\substack{u \in \{u \in \mathcal{U} : r(u) = 0\} \\ (\Delta_\beta, \Delta_L) \in \mathcal{R}_u}} \left(\mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L\right\|_F^2 - \left[(1 - \varepsilon_1)\sigma_{min}^2 \wedge 1\right] \left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right)\right) \geq 0. \tag{S.B.13}$$

Case 2.1. $r(u) \geq 1$ and $\|\mathcal{P}_{\Phi(u)}\Delta_L\|_F \leq \kappa_3\|\Delta_\beta\|_F$. By the definition of $\mathcal{R}_u$, $\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\|_F \leq \sqrt{3r(u)}\kappa_1|\mathcal{P}_{\Phi(u)}\Delta_L\|_F + \kappa_2\|\Delta_\beta\|_F$ if $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$. By the definition of κ_3, we then have

$$\left\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\right\|_F \leq \sqrt{3r(u)}\kappa_1\left(\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F + \kappa_3\|\Delta_\beta\|_F\right) \leq 2\sqrt{3r(u)}\kappa_1\kappa_3\|\Delta_\beta\|_F.$$

Therefore, $\|\Delta_L\|_F^2 \leq (1 + 12r(u)\kappa_1^2)\,\kappa_3^2\|\Delta_\beta\|_F^2$. So, the order of $\|\Delta_L\|_F^2$ is $o\left((N \wedge T))^2\right)\|\Delta_\beta\|_F^2$ uniformly in u, dominated by $NT\|\Delta_\beta\|_F^2$. Therefore, similar to equation (S.B.13), for some N_2 and T_2 that do not change across u, we have the following for all $N > N_2$ and $T > N_2$:

$$\inf_{\substack{u\in\{u\in\mathcal{U}:r(u)\geq 1\}\\(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\\\|\mathcal{P}_{\Phi(u)}\Delta_L\|_F\leq\kappa_3\|\Delta_\beta\|_F}}\left(\mathbb{E}\left\|\sum_{j=1}^{p}X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2 - \left[(1-\varepsilon_1)\sigma_{min}^2 \wedge 1\right]\left(NT\left\|\Delta_\beta\right\|_F^2 + \|\Delta_L\|_F^2\right)\right) \geq 0. \tag{S.B.14}$$

Case 2.2. $r(u) \geq 1$ and $\|\mathcal{P}_{\Phi(u)}\Delta_L\|_F > \kappa_3\|\Delta_\beta\|_F$. Let $M(u) := \sqrt{\mathbb{E}\left\|\mathcal{P}_{\Phi(u)}\left(\sum_{j=1}^p X_j\Delta_{\beta,j}\right)\right\|_F^2}$ and $M(u)^\perp := \sqrt{\mathbb{E}\left\|\mathcal{P}_{\Phi(u)^\perp}\left(\sum_{j=1}^p X_j\Delta_{\beta,j}\right)\right\|_F^2}$. We need the following lemma whose proof is in Appendix S.B.4:

Lemma S.B.3. *Under Assumptions 3 and 4, if $r(u) \geq 1$ for all $u \in \mathcal{U}$, then there exist N_3 and T_3 such that the following holds for all $N > N_3$ and $T > T_3$:*

$$\inf_{u\in\mathcal{U}}\left[M(u)^\perp - C_{\Phi X}\sqrt{3r(u)}\kappa_1\left(M(u) + \kappa_3\left\|\Delta_\beta\right\|_F\right)\right] \geq 0.$$

Remark S.B.2. *Recall that for any $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$, by the definition of the cone, we can verify that the following holds:*

$$\left\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\right\|_F \leq \left(\sqrt{3r(u)}\kappa_1(\lambda)\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F + \kappa_2(\lambda)\left\|\Delta_\beta\right\|_F\right).$$

Comparing the above inequality and Lemma S.B.3, by $C_{\Phi X} > 1$, we can see that our assumptions guarantee that on average, $(\Delta_\beta, \sum_{j=1}^p X_j\Delta_{\beta,j})$ lies away from the cone where (Δ_β, Δ_L) lies in, and the distance is controlled by $C_{\Phi X}$.

We now only consider $N > N_3$ and $T > T_3$ so that the result in Lemma S.B.3 holds. For a fixed $u \in \mathcal{U}$ such that $r(u) \geq 1$, Lemma S.B.3 implies that there exists a $c_1(u) \in \left[0, 1/\left(\sqrt{3r(u)}\kappa_1 C_{\Phi X}\right)\right]$ such that ($c_1(u)$ may depend on Δ_L and Δ_β)

$$c_1(u)M(u)^\perp = M(u) + \kappa_3\|\Delta_\beta\|_F \implies M(u) = c_1(u)M(u)^\perp - \kappa_3\left\|\Delta_\beta\right\|_F, \tag{S.B.15}$$

where $c_1(u) = 0$ if and only if $\Delta_\beta = 0$. Since $\kappa_1 = (2 + C_\lambda)/C_\lambda > 1$, $1 \le r(u) \le \bar{r}$, and $C_{\Phi X} > 1$, we have

$$0 \le c_1(u) \le \frac{1}{\sqrt{3r(u)}\kappa_1 C_{\Phi X}} < \frac{1}{\sqrt{3r(u)}\kappa_1} < 1, \tag{S.B.16}$$

for all $u \in \{u \in \mathcal{U} : r(u) \ge 1\}$ and all (Δ_β, Δ_L).

Similarly, for $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$, since $\left\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\right\|_F \le \sqrt{3r(u)}\kappa_1 \left(\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F + \kappa_3\|\Delta_\beta\|_F\right)$ by the definition of $\mathcal{R}_u$, there exists a $c_2(u) \in \left[0, \sqrt{3r(u)}\kappa_1\right]$ that may depend on Δ_L and Δ_β as well such that

$$\left\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\right\|_F = c_2(u) \left(\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F + \kappa_3 \left\|\Delta_\beta\right\|_F\right). \tag{S.B.17}$$

Then we have

$$0 \le c_2(u) \le \sqrt{3r(u)}\kappa_1 = \frac{\sqrt{3\bar{r}}(2 + C_\lambda)}{C_\lambda}, \tag{S.B.18}$$

$$0 \le c_1(u)c_2(u) \le \frac{1}{C_{\Phi X}} < 1. \tag{S.B.19}$$

for all $u \in \{u \in \mathcal{U} : r(u) \ge 1\}$ and all (Δ_β, Δ_L).

By equations (S.B.15) and (S.B.17), the following inequalities hold for all $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$ and for all $u \in \{u \in \mathcal{U} : r(u) \ge 1\}$:

$$\begin{aligned}
&\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2 \\
=&\mathbb{E}\left\|\mathcal{P}_{\Phi(u)}\left(\sum_{j=1}^p X_j\Delta_{\beta,j}\right) + \mathcal{P}_{\Phi(u)}\Delta_L\right\|_F^2 + \mathbb{E}\left\|\mathcal{P}_{\Phi(u)^\perp}\left(\sum_{j=1}^p X_j\Delta_{\beta,j}\right) + \mathcal{P}_{\Phi(u)^\perp}\Delta_L\right\|_F^2 \\
\ge&\left(M(u) - \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)^2 + \left(M(u)^\perp - \left\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\right\|_F\right)^2 \\
=&\left(c_1(u)M(u)^\perp - \left(\kappa_3\left\|\Delta_\beta\right\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)\right)^2 + \left(M(u)^\perp - c_2(u)\left(\kappa_3\left\|\Delta_\beta\right\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)\right)^2.
\end{aligned} \tag{S.B.20}$$

The inequality is by equation (S.B.12). The last equality is by equations (S.B.15) and (S.B.17). Since $c_1(u)c_2(u)$ is bounded away from 1 from above (equation (S.B.19)), the two squared terms on the right hand side of equation (S.B.20) cannot be simultaneously 0 for any $u \in \{u \in \mathcal{U} : r(u) \ge 1\}$ unless that all the norms in them are zero. We now derive a lower bound on the right hand side of equation (S.B.20).

By equation (S.B.15) and by the uniform boundedness of $c_1(u)$, there exists a constant

$C_1 > 1$ not depending on u or (Δ_β, Δ_L) such that

$$C_1 \left(M(u)^\perp\right)^2 \geq \left(1 + c_1(u)^2\right) \left(M(u)^\perp\right)^2 \geq \left(M(u)^\perp\right)^2 + (M(u))^2 = \mathbb{E}\left\|\sum_{j=1}^p X_j \Delta_{\beta,j}\right\|_F^2. \tag{S.B.21}$$

Similarly, by equation (S.B.17), by $\|\mathcal{P}_{\Phi(u)}\Delta_L\|_F > \kappa_3\|\Delta_\beta\|_F$, and by the uniform boundedness of $c_2(u)$ (by equation (S.B.18)), we have some $C_2 > 1$ (C_2 decreases in C_λ) that does not depend on u or (Δ_β, Δ_L) such that

$$\|\Delta_L\|_F^2 = \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F^2 + \left\|\mathcal{P}_{\Phi(u)^\perp}\Delta_L\right\|_F^2 < \left(1 + 4c_2(u)^2\right)\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F^2 \leq C_2\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F^2. \tag{S.B.22}$$

Hence, we have the following holding for all $u \in \{u \in \mathcal{U} : r(u) \geq 1\}$ and $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$ such that $\left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F - \kappa_3\|\Delta_\beta\|_F > 0$,

$$\begin{aligned}
&\left[c_1(u)M(u)^\perp - \left(\kappa_3\|\Delta_\beta\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)\right]^2 + \left[M(u)^\perp - c_2(u)\left(\kappa_3\|\Delta_\beta\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)\right]^2 \\
&= \left(1 + c_1(u)^2\right)\left(M(u)^\perp\right)^2 + \left(1 + c_2(u)^2\right)\left(\kappa_3\|\Delta_\beta\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)^2 \\
&\quad - 2\left(c_1(u) + c_2(u)\right)M(u)^\perp\left(\kappa_3\|\Delta_\beta\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right) \\
&\geq \left(1 - \frac{c_1(u) + c_2(u)}{\sqrt{(c_1(u) + c_2(u))^2 + (1 - c_1(u)c_2(u))^2}}\right)\left[\left(M(u)^\perp\right)^2 + \left(\kappa_3\|\Delta_\beta\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)^2\right] \\
&\geq \left(1 - \frac{\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)}{\sqrt{\left(\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)\right)^2 + (1 - 1/C_{\Phi X})^2}}\right)\left[\left(M(u)^\perp\right)^2 + \left(\kappa_3\|\Delta_\beta\|_F + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F\right)^2\right] \\
&\geq \left(1 - \frac{\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)}{\sqrt{\left(\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)\right)^2 + (1 - 1/C_{\Phi X})^2}}\right)\left[\left(M(u)^\perp\right)^2 + \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F^2\right] \\
&\geq \left(1 - \frac{\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)}{\sqrt{\left(\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)\right)^2 + (1 - 1/C_{\Phi X})^2}}\right)\left(\frac{1}{C_1}\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j}\right\|_F^2 + \frac{1}{C_2}\|\Delta_L\|_F^2\right) \\
&\geq \left(\frac{\sigma_{min}^2}{C_1} \wedge \frac{1}{C_2}\right)\left(1 - \frac{\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)}{\sqrt{\left(\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)\right)^2 + (1 - 1/C_{\Phi X})^2}}\right)\left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right) \\
&\geq \left(\frac{\sigma_{min}^2}{C_1} \wedge \frac{1}{C_2}\right)\left(1 - \frac{\sqrt{3\bar{r}}\kappa_1 + 1/(\sqrt{3\bar{r}}\kappa_1)}{\sqrt{\left(\sqrt{3\bar{r}}\kappa_1 + 1/(\sqrt{3\bar{r}}\kappa_1)\right)^2 + (1 - 1/C_{\Phi X})^2}}\right)\left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right)
\end{aligned}$$

where the equality is elementary. The first inequality is due to the fact that for any nonnegative real numbers k_1 and k_2, and real numbers x and y, the following holds:[5]

$$(1+k_1^2)x^2+(1+k_2^2)y^2-2(k_1+k_2)xy \geq \left(1-\frac{k_1+k_2}{\sqrt{(k_1+k_2)^2+(1-k_1k_2)^2}}\right)(x^2+y^2), \quad \text{(S.B.23)}$$

where the coefficient on (x^2+y^2) on the right hand side is strictly positive as long as $k_1k_2 \neq 1$. The second inequality is by equations (S.B.16), (S.B.18) and (S.B.19). The third inequality is elementary. The fourth inequality is by equations (S.B.21) and (S.B.22). The fifth inequality is by equation (S.B.10). The last inequality holds because of the following reason: Since $\kappa_1 = (2+C_\lambda)/C_\lambda$, $\sqrt{3r(u)}\kappa_1 > 1$ by construction. Therefore, $\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1)$ increases as $\sqrt{3r(u)}\kappa_1$ increases. So $\sqrt{3r(u)}\kappa_1 + 1/(\sqrt{3r(u)}\kappa_1) \leq \sqrt{3\bar{r}}\kappa_1 + 1/(\sqrt{3\bar{r}}\kappa_1)$. Let

$$C_{RSC} \coloneqq \left(\frac{\sigma_{min}^2}{C_1} \wedge \frac{1}{C_2}\right)\left(1-\frac{\sqrt{3\bar{r}}\kappa_1 + 1/(\sqrt{3\bar{r}}\kappa_1)}{\sqrt{\left(\sqrt{3\bar{r}}\kappa_1 + 1/(\sqrt{3\bar{r}}\kappa_1)\right)^2 + (1-1/C_{\Phi X})^2}}\right). \quad \text{(S.B.24)}$$

Again, since $\sqrt{3\bar{r}}\kappa_1 > 1$, C_{RSC} gets bigger when κ_1 and C_2 get smaller, which is a consequence of a bigger C_λ. Meanwhile, C_{RSC} is also increasing in $C_{\Phi X}$. The above derivation has shown that for $N > N_3$ and $T > T_3$ so that Lemma S.B.3 holds,

$$\inf_{\substack{u\in\{u\in\mathcal{U}:r(u)\geq 1\} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u \\ \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F - \kappa_3\|\Delta_\beta\|_F > 0}} \left(\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2 - C_{RSC}\left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right)\right) \geq 0. \quad \text{(S.B.25)}$$

Comparing equations (S.B.13), (S.B.14) and (S.B.25), since ε_1 can be arbitrarily small and $C_1, C_2 > 1$, we have $C_{RSC} < [(1-\varepsilon_1)\sigma_{min}^2 \wedge 1]$. Therefore, for all $N > \max\{N_1, N_2, N_3\}$ and $T > \max\{T_1, T_2, T_3\}$, we have

$$\begin{aligned}
&\inf_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u}} \left(\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2 - C_{RSC}\left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right)\right) \\
&= \min\left\{\inf_{\substack{u\in\{u\in\mathcal{U}:r(u)=0\} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u}} \left(\mathbb{E}\left\|\sum_{j=1}^p X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2 - C_{RSC}\left(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2\right)\right),\right.
\end{aligned}$$

[5] To see this, let $\psi \coloneqq \frac{k_1+k_2}{\sqrt{(k_1+k_2)^2+(1-k_1k_2)^2}}$. We can see $\psi \in [0,1]$. Then $\psi(1+k_1^2)x^2 + \psi(1+k_2^2)y^2 - 2(k_1+k_2)xy \geq 0$ because $(k_1+k_2)^2 = \psi^2(1+k_1^2)(1+k_2^2)$. Therefore, $(1+k_1^2)x^2 + (1+k_2^2)y^2 - 2(k_1+k_2)xy \geq (1-\psi)((1+k_1^2)x^2 + (1+k_2^2)y^2) \geq (1-\psi)(x^2+y^2)$.

$$
\inf_{\substack{u\in\{u\in\mathcal{U}:r(u)\geq 1\}\\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\\ \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F-\kappa_3\|\Delta_\beta\|_F\leq 0}} \left(\mathbb{E}\left\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}+\Delta_L\right\|_F^2 - C_{RSC}\left(NT\left\|\Delta_\beta\right\|_F^2+\left\|\Delta_L\right\|_F^2\right)\right),
$$

$$
\left.\inf_{\substack{u\in\{u\in\mathcal{U}:r(u)\geq 1\}\\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\\ \left\|\mathcal{P}_{\Phi(u)}\Delta_L\right\|_F-\kappa_3\|\Delta_\beta\|_F> 0}} \left(\mathbb{E}\left\|\sum_{j=1}^{p} X_j\Delta_{\beta,j}+\Delta_L\right\|_F^2 - C_{RSC}\left(NT\left\|\Delta_\beta\right\|_F^2+\left\|\Delta_L\right\|_F^2\right)\right)\right\}
$$

$$
\geq 0.
$$

Finally, since Lemma S.B.2 shows $\varepsilon_{NT} \to 0$, for $N > N_4$ and $T > T_4$ for some N_4 and T_4, we have $\varepsilon_{NT} < C_{RSC}/2$. The desired result in Theorem 1 obtains for all $N > \max\{N_1, N_2, N_3, N_4\}$ and $T > \max\{T_1, T_2, T_3, T_4\}$ by substituting the above inequality and $\varepsilon_{NT} < C_{RSC}/2$ into equation (S.B.8) in Lemma S.B.2. The rank of $L_0(u)$ is identified immediately once $L_0(u)$ is identified. ■

S.B.3 Proofs of the Results in Section 5

Proof of Theorem 2

For arbitrary random variables Z_{it}s and a function f, let $\mathbb{G}_u(f(Z_{it})) \coloneqq \sum_{i,t}[f(Z_{it}) - \mathbb{E}(f(Z_{it}))]/\sqrt{NT}$. The proof of the following lemma is in Appendix S.B.4.

Lemma S.B.4. *Under Assumptions 1, 3 and 5, there exists a constant $C_{sup} > 0$ such that for γ defined in equation (3.4),*

$$
\mathbb{P}\left(\sup_{\substack{u\in\mathcal{U}\\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2\leq\gamma^2}} \left|\mathbb{G}_u\left(\rho_u\left(V_{it}(u)-X_{it}'\Delta_\beta-\Delta_{L,it}\right)-\rho_u(V_{it}(u))\right)\right| \leq C_{sup}\sqrt{\log(NT)}\left(\sqrt{p\log((p+1)NT)}\vee\sqrt{\bar{r}(N\vee T)}\right)\gamma\right) \to 1.
$$

The positive constant C_{sup} decreases in C_λ and is lower bounded by a positive constant. Its formula is in the proof.

Let Ω_0 be the event that i) $(\hat{\Delta}_\beta(u), \hat{\Delta}_L(u)) \in \mathcal{R}_u \cap \mathcal{D}$ and ii) the uniform bound on the error process $\mathbb{G}_u$ in Lemma S.B.4 holds. By Lemmas 1 and S.B.4 and by equation (3.1) under Ω_L, the event Ω_0 occurs w.p.a.1. It is then sufficient to show that the following event has zero probability under Ω_0:

$$\exists u \in \mathcal{U} : \|\hat{\Delta}_\beta(u)\|_F^2 + \frac{1}{NT}\|\hat{\Delta}_L(u)\|_F^2 > \gamma^2 \tag{S.B.26}$$

where $\gamma = C_{error}\alpha_{NT}^2 \left((1+C_\lambda) \vee \sqrt{\log(NT)}\right)\left(\sqrt{p\log((p+1)NT)/NT} \vee \sqrt{\bar{r}/(N \wedge T)}\right)$ for some $C_{error} > 0$ about which we will be precise later.

Since $\mathcal{R}_u$ is a cone and zero is contained in $\mathcal{D}$ which is a convex set, for any $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u \cap \mathcal{D}$ and any $\tau \in (0,1)$, $(\tau\Delta_\beta, \tau\Delta_L) \in \mathcal{R}_u \cap \mathcal{D}$. By this observation, by the definition of the estimator (2.3) and $L_0(u) \in \mathcal{L}$ for all $u \in \mathcal{U}$, and by convexity of the objective function, equation (S.B.26) implies that there exists a $u \in \mathcal{U}$ such that

$$\begin{aligned}
0 \geq & \inf_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2=\gamma^2}} \frac{1}{NT}\left[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j\Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u\left(V(u)\right)\right] \\
& \qquad + \lambda\left[\|L_0(u)+\Delta_L\|_* - \|L_0(u)\|_*\right] \qquad \text{(S.B.27)} \\
= & \inf_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2=\gamma^2}} \frac{1}{NT}\mathbb{E}\left[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j\Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\right] \\
& \qquad + \frac{1}{\sqrt{NT}}\mathbb{G}_u\left(\rho_u\Big(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\Big) - \rho_u(V_{it}(u))\right) \\
& \qquad + \lambda\left[\|L_0(u)+\Delta_L\|_* - \|L_0(u)\|_*\right]. \qquad \text{(S.B.28)}
\end{aligned}$$

For the expectation, Theorem 1 implies that

$$\begin{aligned}
& \inf_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2=\gamma^2}} \frac{1}{NT}\mathbb{E}\left[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j\Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\right] \\
\geq & \frac{C_{min}C_{RSC}}{2\alpha_{NT}^2}\left(\|\Delta_\beta\|_F^2 + \frac{\|\Delta_L\|_F^2}{NT}\right) \\
= & \frac{C_{min}C_{RSC}}{2\alpha_{NT}^2}\gamma^2, \qquad \text{(S.B.29)}
\end{aligned}$$

where $C_{min} = (1 \wedge \delta)^2\underline{f}/18$.

For the error process $\mathbb{G}_u/\sqrt{NT}$, by ii) in Ω_0,

$$\frac{1}{\sqrt{NT}} \sup_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u \\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2=\gamma^2}} \left|\mathbb{G}_u\left(\rho_u\Big(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\Big) - \rho_u(V_{it}(u))\right)\right|$$

$$\leq C_{sup}\sqrt{\log(NT)}\left(\sqrt{\frac{p\log((p+1)NT)}{NT}}\vee\sqrt{\frac{\bar{r}}{N\wedge T}}\right)\gamma. \tag{S.B.30}$$

Finally, for the penalty difference, same as the proof of Lemma S.B.4, we define $C_{Cone} = 5\sqrt{2C_X}/(C_\lambda C_{op})$. By $\lambda = (1+C_\lambda)C_{op}\sqrt{N\vee T}/NT$, and by the definition of $\mathcal{R}_u$,

$$\begin{aligned}
&\sup_{\substack{u\in\mathcal{U}\\(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2=\gamma^2}} \lambda\Big|\|L_0(u)+\Delta_L\|_*-\|L_0(u)\|_*\Big|\\
\leq&\sup_{\substack{u\in\mathcal{U}\\(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2=\gamma^2}} \lambda\|\Delta_L\|_*\\
\leq&\lambda\sup_{\|\Delta_L\|_F^2\leq NT\gamma^2}\left(C_{Cone}\sqrt{p(N\wedge T)\log((p+1)NT)}\gamma+\frac{2(1+C_\lambda)}{C_\lambda}\|\mathcal{P}_{\Phi(u)}\Delta_L\|_*\right)\\
\leq&\lambda\sup_{\|\Delta_L\|_F^2\leq NT\gamma^2}\left(C_{Cone}\sqrt{p(N\wedge T)\log((p+1)NT)}\gamma+\frac{2\sqrt{3\bar{r}}(1+C_\lambda)}{C_\lambda}\|\Delta_L\|_F\right)\\
\leq&(1+C_\lambda)C_{op}\left(C_{Cone}+\frac{2\sqrt{3}(1+C_\lambda)}{C_\lambda}\right)\left(\sqrt{\frac{p\log((p+1)NT)}{NT}}\vee\sqrt{\frac{\bar{r}}{N\wedge T}}\right)\gamma\\
\leq&(1+C_\lambda)C_{pel}\left(\sqrt{\frac{p\log((p+1)NT)}{NT}}\vee\sqrt{\frac{\bar{r}}{N\wedge T}}\right)\gamma,
\end{aligned} \tag{S.B.31}$$

where $C_{pel} := C_{op}\left(C_{Cone}+2\sqrt{3}(1+C_\lambda)/C_\lambda\right)$. We can see that C_{pel} decreases as C_λ increases, and is bounded away from zero.

Let $C_{up} = C_{pel} + C_{sup}$, so C_{up} is bounded away from zero and decreases as C_λ increases as well. Let $C_{error} = 3C_{up}/(C_{min}C_{RSC})$. By equations (S.B.29), (S.B.30) and (S.B.31), the right side of equation (S.B.28) is lower bounded by

$$\gamma\cdot\left(\frac{C_{min}C_{RSC}}{2\alpha_{NT}^2}\gamma - C_{up}\left((1+C_\lambda)\vee\sqrt{\log(NT)}\right)\left(\sqrt{\frac{p\log((p+1)NT)}{NT}}\vee\sqrt{\frac{\bar{r}}{N\wedge T}}\right)\right)>0,$$

with probability one. Hence, under Ω_0, inequality (S.B.27) and thus inequality (S.B.26) hold with zero probability. Since Ω_0 holds w.p.a.1 (implicitly conditional on (W_L,Ω_L)), we obtain the desired result. ■

Remark S.B.3. *As noted in Section 3.1, all the arguments are implicitly conditional on (W_L,Ω_L). So here our result, driven by $\mathbb{P}(\Omega_0|W_L,\Omega_L)\to 1$, is also conditional on them. However, under the assumption that $\mathbb{P}(\Omega_L)\to 1$, the results also hold unconditionally as $\mathbb{P}(\Omega_0)\geq\mathbb{E}(\mathbb{P}(\Omega_0|\Omega_L,W_L)|\Omega_L)\mathbb{P}(\Omega_L)\to 1$.*

Proof of Corollary 1

By Weyl's inequality for singular values,

$$\max_{k\in\{1,\dots,N\wedge T\}}\{|\hat{\sigma}_k(u)-\sigma_k(u)|\}\le\|\hat{\Delta}_L(u)\|\le\|\hat{\Delta}_L(u)\|_F,$$

with probability one. Hence, $\sup_{u\in\mathcal{U}}\max_k\{|\hat{\sigma}_k(u)-\sigma_k(u)|\}\le\sup_{u\in\mathcal{U}}\|\hat{\Delta}_L(u)\|_F$. Equation (5.3) thus follows by plugging in the uniform rate of $\|\hat{\Delta}_L(u)\|_F$ obtained in Theorem 2 and by $\sigma_{r(u)+1}(u)=\cdots=\sigma_{N\wedge T}(u)=0$. ∎

Proof of Corollary 2

By the definition of $\hat{r}(u)$, the event $\{\hat{r}(u)=r(u)\}$ is equivalent to $\{\hat{\sigma}_{r(u)}\ge C_r\}\cap\{\hat{\sigma}_{r(u)+1}<C_r\}$. The latter event, under the event $\Omega_{sv2}=\{\sigma_{r(u)}$ is of the order of $\sqrt{NT}\}$, can be implied by $\Omega_{sv1}:=\{|\hat{\sigma}_{r(u)}-\sigma_{r(u)}|\le\sqrt{NT}\gamma,|\hat{\sigma}_{r(u)+1}-0|\le\sqrt{NT}\gamma\}$ for sufficiently large N and T by the choice of C_r. The desired result is thus obtained since w.p.a.1, Ω_{sv1} is true by Corollary 1 and Ω_{sv2} is true by assumption.

S.B.4 Proofs of the Results in Appendices S.A and S.B

Proof of Lemma S.B.1

Let Ω_1 be the event that $\max_{1\le j\le p}\|X_j\|_F^2\le C_XNT$. Under Assumption 1, $\mathbb{P}(\Omega_1)\to 1$. Recall that the (i,t)-th element in subgradient $\nabla\boldsymbol{\rho}_u(V(u))$ is

$$\left(\nabla\boldsymbol{\rho}_u\left(V(u)\right)\right)_{it}=u\mathbb{1}(V_{it}(u)\ge 0)+(u-1)\mathbb{1}(V_{it}(u)<0),$$

with probability one. By Assumption 1 and by $V(u)=Y-q_{Y|W_X}(u)$, the elements in $\nabla\boldsymbol{\rho}_u\left(V(u)\right)$ are independent with mean 0 conditional on W_X, and are uniformly bounded within $[-1,1]$. We start by proving equation (S.B.1).

Proof of Equation (S.B.1). Let $M=5\sqrt{2C_XNT\log((p+1)NT)}$. Note that

$$\mathbb{P}\left(\sup_{\substack{u\in\mathcal{U}\\1\le j\le p}}|\langle\nabla\boldsymbol{\rho}_u(V(u)),X_j\rangle|>M\right)\le\mathbb{P}\left(\sup_{\substack{u\in\mathcal{U}\\1\le j\le p}}|\langle\nabla\boldsymbol{\rho}_u(V(u)),X_j\rangle|>M\middle|\Omega_1\right)\mathbb{P}(\Omega_1)+\mathbb{P}(\Omega_1^c).\tag{S.B.32}$$

Since $\mathbb{P}(\Omega_1^c)\to 0$, it is sufficient to show the conditional probability in equation (S.B.32) converges to zero.

Let $\mathcal{U}_K=(u_1,u_2,\dots,u_K)$ be an ε-net of $\mathcal{U}$. Let $\varepsilon=\frac{1}{\sqrt{NT}}$ and $K\varepsilon\le 1$. By the triangle inequality,

$$
\begin{aligned}
&\sup_{\substack{u\in\mathcal{U}\\1\le j\le p}}\left|\left\langle\nabla\boldsymbol{\rho}_u(V(u)),X_j\right\rangle\right|\\
\le&\max_{\substack{u_k\in\mathcal{U}_K\\1\le j\le p}}\left|\left\langle\nabla\boldsymbol{\rho}_{u_k}(V(u_k)),X_j\right\rangle\right|+\sup_{\substack{|u-u_k|\le\varepsilon,u_k\in\mathcal{U}_K\\1\le j\le p}}\left|\left\langle\nabla\boldsymbol{\rho}_u(V(u))-\nabla\boldsymbol{\rho}_{u_k}(V(u_k)),X_j\right\rangle\right|\\
=:&\mathcal{E}_1+\mathcal{E}_2.
\end{aligned}
$$

Then by the union bound,

$$
\mathbb{P}\left(\sup_{\substack{u\in\mathcal{U}\\1\le j\le p}}|\langle\nabla\boldsymbol{\rho}_u(V(u)),X_j\rangle|>M\middle|\Omega_1\right)\le2\max\left\{\mathbb{P}\left(\mathcal{E}_1>\frac{1}{5}M\middle|\Omega_1\right),\mathbb{P}\left(\mathcal{E}_2>\frac{4}{5}M\middle|\Omega_1\right)\right\}. \tag{S.B.33}
$$

Bound on $\mathcal{E}_1$. By $K\le1/\varepsilon$, we have

$$
\begin{aligned}
\mathbb{P}\left(\mathcal{E}_1>\frac{1}{5}M\middle|\Omega_1\right)\le&\frac{p}{\varepsilon}\max_{\substack{u_k\in\mathcal{U}_K\\1\le j\le p}}\mathbb{P}\left(\left|\left\langle\nabla\boldsymbol{\rho}_{u_k}(V(u_k)),X_j\right\rangle\right|\ge\frac{1}{5}M\middle|\Omega_1\right)\\
=&\frac{p}{\varepsilon}\max_{\substack{u_k\in\mathcal{U}_K\\1\le j\le p}}\mathbb{E}\left[\mathbb{P}\left(\left|\left\langle\nabla\boldsymbol{\rho}_{u_k}(V(u_k)),X_j\right\rangle\right|\ge\frac{1}{5}M\middle|\Omega_1,W_X\right)\middle|\Omega_1\right]\\
\le&2p\sqrt{NT}\exp\left(-\frac{M^2}{50C_XNT}\right)\\
=&\frac{2p}{(p+1)\sqrt{NT}}\to0
\end{aligned}\tag{S.B.34}
$$

where the first inequality is by the union bound and the following equality is due to the law of iterative expectation. The penultimate inequality is by Hoeffding's inequality and by Ω_1 under $\varepsilon=1/\sqrt{NT}$.

Bound on $\mathcal{E}_2$. By definition, the (i,t)-th element in $\nabla\boldsymbol{\rho}_u(V(u))-\nabla\boldsymbol{\rho}_{u_k}(V(u_k))$ is almost surely

$$
\begin{aligned}
&u\mathbb{1}(V_{it}(u)\ge0)+(u-1)\mathbb{1}(V_{it}(u)<0)-[u_k\mathbb{1}(V_{it}(u_k)\ge0)+(u_k-1)\mathbb{1}(V_{it}(u_k)<0)]\\
=&(u-u_k)+\mathbb{1}(V_{it}(u_k)<0)-\mathbb{1}(V_{it}(u)<0).
\end{aligned}\tag{S.B.35}
$$

Let Ξ_1 and Ξ_2 be two $N\times T$ matrices whose (i,t)-th elements are

$$
\Xi_{1,it}\coloneqq u-u_k \tag{S.B.36}
$$

$$
\Xi_{2,it}\coloneqq\mathbb{1}(V_{it}(u_k)<0)-\mathbb{1}(V_{it}(u)<0). \tag{S.B.37}
$$

Then by equation (S.B.35), $\nabla\boldsymbol{\rho}_u(V(u)) - \nabla\boldsymbol{\rho}_{u_k}(V(u_k)) = \Xi_1 + \Xi_2$. Therefore,

$$\mathbb{P}\left(\mathcal{E}_2 > \frac{4}{5}M\middle|\Omega_1\right) \leq \mathbb{P}\left(\sup_{\substack{|u-u_k|\leq\varepsilon\\ u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} |\langle\Xi_1, X_j\rangle| > \frac{2}{5}M\middle|\Omega_1\right) + \mathbb{P}\left(\sup_{\substack{|u-u_k|\leq\varepsilon\\ u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} |\langle\Xi_2, X_j\rangle| > \frac{2}{5}M\middle|\Omega_1\right)$$
$$=:\mathbb{P}_1 + \mathbb{P}_2. \tag{S.B.38}$$

We first show $\mathbb{P}_1 = 0$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{P}\left(\sup_{\substack{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} |\langle\Xi_1, X_j\rangle| > \frac{2}{5}M\middle|\Omega_1\right) &\leq \mathbb{P}\left(\sup_{\substack{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} \|\Xi_1\|_F\|X_j\|_F > \frac{2}{5}M\middle|\Omega_1\right)\\ &\leq \mathbb{P}\left(\varepsilon\sqrt{C_X}NT > \frac{2}{5}M\right)\\ &= \mathbb{P}\left(\sqrt{C_XNT} > 2\sqrt{2C_XNT\log((p+1)NT)}\right)\\ &= 0\end{aligned} \tag{S.B.39}$$

for large enough N and T. The second inequality is by the definition of Ξ_1 and Ω_1. The penultimate equality is by $\varepsilon = 1/\sqrt{NT}$.

Now we show that $\mathbb{P}_2$ converges to zero. Let $\Xi_2^{(1)}$ and $\Xi_2^{(2)}$ be two $N\times T$ matrices whose (i,t)-th elements are

$$\Xi_{2,it}^{(1)}(u_k) := \mathbb{1}(V_{it}(u_k) < 0) - \mathbb{1}(V_{it}(u_k - \varepsilon) < 0). \tag{S.B.40}$$
$$\Xi_{2,it}^{(2)}(u_k) := \mathbb{1}(V_{it}(u_k) < 0) - \mathbb{1}(V_{it}(u_k + \varepsilon) < 0). \tag{S.B.41}$$

Consider an arbitrary element in Ξ_2: $\Xi_{2,it} := \mathbb{1}(V_{it}(u_k) < 0) - \mathbb{1}(V_{it}(u) < 0)$. By Assumption 1, $V_{it}(u)$ is strictly decreasing in u almost surely. Hence, $\mathbb{1}(V_{it}(u) < 0)$ is weakly increasing in u almost surely. Consequently, if $u_k - \varepsilon \leq u \leq u_k$, then $0 \leq \Xi_{2,it} \leq \Xi_{2,it}^{(1)}(u_k) \leq 1$. Similarly, if $u_k + \varepsilon \geq u \geq u_k$, then $0 \geq \Xi_{2,it} \geq \Xi_{2,it}^{(2)}(u_k) \geq -1$. The following inequalities thus hold with probability one.

$$\begin{aligned}\sup_{\substack{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} \left|\langle\Xi_2, X_j\rangle\right| &\leq \sup_{\substack{u_k-\varepsilon\leq u\leq u_k, u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} \left|\langle\Xi_2, X_j\rangle\right| + \sup_{\substack{u_k\leq u\leq u_k+\varepsilon, u_k\in\mathcal{U}_k\\ 1\leq j\leq p}} \left|\langle\Xi_2, X_j\rangle\right|\\ &\leq \sup_{\substack{u_k-\varepsilon\leq u\leq u_k, u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} \left|\langle\Xi_2, |X_j|\rangle\right| + \sup_{\substack{u_k\leq u\leq u_k+\varepsilon, u_k\in\mathcal{U}_k\\ 1\leq j\leq p}} \left|\langle\Xi_2, |X_j|\rangle\right|\\ &\leq \max_{\substack{u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} \left|\langle\Xi_2^{(1)}, |X_j|\rangle\right| + \max_{\substack{u_k\in\mathcal{U}_K\\ 1\leq j\leq p}} \left|\langle\Xi_2^{(2)}, |X_j|\rangle\right|.\end{aligned}$$

The first inequality is elementary. To see why the second inequality holds, note that the elements in Ξ_2 are all nonnegative when $u \in [u_k - \varepsilon, u_k]$ and are all nonpositive when $u \in [u_k, u_k + \varepsilon]$. So, for a given Ξ_2 and a given X_j, the two absolute inner products on the right side of the first inequality increase if we flip the signs of the $X_{j,it}$s so that they also have the same sign. The third inequality then follows because now that elements in both $|X_j|$ and in Ξ_2 have the same signs, the two absolute inner products in the second line increase as the magnitude of any of the elements in Ξ_2 increases. Therefore,

$$\mathbb{P}_2 \leq \mathbb{P}\left(\max_{\substack{u_k \in \mathcal{U}_K \\ 1\leq j\leq p}} \left|\langle \Xi_2^{(1)}, |X_j|\rangle\right| > \frac{1}{5}M \middle| \Omega_1\right) + \mathbb{P}\left(\max_{\substack{u_k \in \mathcal{U}_K \\ 1\leq j\leq p}} \left|\langle \Xi_2^{(2)}, |X_j|\rangle\right| > \frac{1}{5}M \middle| \Omega_1\right). \tag{S.B.42}$$

Let us first bound $\max_{u_k\in\mathcal{U}_k, 1\leq j\leq p} \left|\langle \Xi_2^{(1)}, |X_j|\rangle\right|$. The expectation of an arbitrary element $\Xi_{2,it}^{(1)}$ in $\Xi_2^{(1)}$ satisfies

$$\begin{aligned}
\mathbb{E}\left(\Xi_{2,it}^{(1)}(u_k)\middle| W_X\right) =& \mathbb{P}\left(V_{it}(u_k) < 0 \leq V_{it}(u_k - \varepsilon)\middle| W_X\right) \\
=& \mathbb{P}\left(q_{Y_{it}|W_X}(u_k - \varepsilon) \leq Y_{it} < q_{Y_{it}|W_X}(u_k)\middle| W_X\right) \\
=& \varepsilon
\end{aligned}$$

where the second equality is by the definition of $V_{it}(u)$. Let $\bar{\Xi}_2^{(1)} = \mathbb{E}\left(\Xi_2^{(1)}(u_k)\middle| W_X\right)$ be an $N\times T$ matrix whose elements are all equal to ε. Under Ω_1, by the Cauchy-Schwarz inequality and by $\varepsilon = 1/\sqrt{NT}$, we have $\max_{u_k\in\mathcal{U}_k, 1\leq j\leq p} \left|\langle \bar{\Xi}_2^{(1)}, |X_j|\rangle\right| \leq \varepsilon\sqrt{C_X}NT = \sqrt{C_X NT}$ with probability one. Therefore,

$$\begin{aligned}
&\mathbb{P}\left(\max_{\substack{u_k \in \mathcal{U}_k \\ 1\leq j\leq p}} \left|\langle \Xi_2^{(1)}, |X_j|\rangle\right| > \frac{1}{5}M \middle| \Omega_1\right) \\
\leq& \mathbb{P}\left(\max_{\substack{u_k \in \mathcal{U}_k \\ 1\leq j\leq p}} \left|\langle \Xi_2^{(1)} - \bar{\Xi}_2^{(1)}, |X_j|\rangle\right| + \max_{\substack{u_k \in \mathcal{U}_k \\ 1\leq j\leq p}} \left|\langle \bar{\Xi}_2^{(1)}, |X_j|\rangle\right| > \frac{1}{5}M \middle| \Omega_1\right) \\
\leq& \mathbb{P}\left(\max_{\substack{u_k \in \mathcal{U}_k \\ 1\leq j\leq p}} \left|\langle \Xi_2^{(1)} - \bar{\Xi}_2^{(1)}, |X_j|\rangle\right| > \frac{1}{5}M - \sqrt{C_X NT} \middle| \Omega_1\right) \\
=& \frac{p}{\varepsilon}\max_{\substack{u_k \in \mathcal{U}_k \\ 1\leq j\leq p}} \mathbb{E}\left[\mathbb{P}\left(\left|\langle \Xi_2^{(1)} - \bar{\Xi}_2^{(1)}, |X_j|\rangle\right| > \frac{1}{5}M - \sqrt{C_X NT} \middle| W_X\right)\middle| \Omega_1\right] \\
\leq& 2p\sqrt{NT}\exp\left(-\frac{(M/5 - \sqrt{C_X NT})^2}{2C_X NT}\right) \to 0.
\end{aligned} \tag{S.B.43}$$

The penultimate equality is by the law of iterated expectation since $\max_{1\leq j\leq p} \|X_j\|_F^2$ in Ω_1 is a function of W_X. The last inequality is by Hoeffding's inequality since conditional on W_X, elements in $(\Xi_2^{(1)} - \bar{\Xi}_2^{(1)})$ are independent with zero mean and are bounded within $[-1, 1]$. Convergence is by the choice of M.

Finally, we can show that $\mathbb{P}\left(\max_{u_k \in \mathcal{U}_k, 1 \le j \le p} |\langle \Xi_2^{(2)}, |X_j| \rangle| > M/5 | \Omega_1\right) \to 0$ as well following exactly the same argument. Combining it with equations (S.B.32), (S.B.33), (S.B.34), (S.B.38), (S.B.39), (S.B.42) and (S.B.43), we obtain the desired result.

Proof of Equation (S.B.2). We invoke the following lemma to prove equation (S.B.2).

Lemma S.B.5 (Theorem 4.4.5 in Vershynin (2018), p.85)**.** *Let A be an $N \times T$ random matrix whose entries A_{ij} are independent mean zero sub-Gaussian random variables. Then, for any $t > 0$ we have $\|A\| \le CQ\left(\sqrt{N} + \sqrt{T} + t\right)$ with probability at least $1 - 2\exp(-t^2)$. Here C is an absolute constant and $Q = \max_{i,j} \|A_{ij}\|_{\psi_2}$*[6].

For any u, entries in matrix $\nabla \boldsymbol{\rho}_u(V(u))$ are independent mean zero random variables conditional on W_X. Since each of them is Bernoulli, they are sub-Gaussian with $\|(\nabla \boldsymbol{\rho}_u(V(u)))_{i,t}\|_{\psi_2} = \inf\{c > 0 : (1-u)\exp(u^2/c^2) + u\exp\left((1-u)^2/c^2\right) \le 2\}$. Take $Q = \sup_{u \in (0,1)} \|(\nabla \boldsymbol{\rho}_u(V(u)))_{i,t}\|_{\psi_2} < \infty$. Let $C_{op} = 12CQ + 6$. Let $M = C_{op}\sqrt{N \vee T}$.

Let $\mathcal{U}_K = (u_1, u_2, ..., u_K)$ be an ε-net of $\mathcal{U}$ with $\varepsilon K \le 1$. This time let $\varepsilon = 1/\sqrt{N \vee T}$. By the triangle inequality, we have

$$\begin{aligned}
\sup_{u \in \mathcal{U}} \|\nabla \boldsymbol{\rho}_u(V(u))\| &\le \max_{u_k \in \mathcal{U}_K} \|\nabla \boldsymbol{\rho}_{u_k}(V(u_k))\| + \sup_{|u-u_k| \le \varepsilon, u_k \in \mathcal{U}_K} \|\nabla \boldsymbol{\rho}_u(V(u)) - \nabla \boldsymbol{\rho}_{u_k}(V(u_k))\| \\
&=: \mathcal{F}_1 + \mathcal{F}_2. \qquad \text{(S.B.44)}
\end{aligned}$$

By equation (S.B.44),

$$\mathbb{P}\left(\sup_{u \in \mathcal{U}} \|\nabla \boldsymbol{\rho}_u(V(u))\| > M\right) \le 2\max\left\{\mathbb{P}\left(\mathcal{F}_1 > M/2\right), \mathbb{P}\left(\mathcal{F}_2 > M/2\right)\right\}. \qquad \text{(S.B.45)}$$

Bound on $\mathcal{F}_1$. By $\varepsilon K \le 1$, we have

$$\begin{aligned}
\mathbb{P}\left(\mathcal{F}_1 > M/2\right) &\le \frac{1}{\varepsilon} \max_{u_k \in \mathcal{U}_K} \mathbb{P}\left(\|\nabla \boldsymbol{\rho}_{u_k}(V(u_k))\| > M/2\right) \\
&= \frac{1}{\varepsilon} \max_{u_k \in \mathcal{U}_K} \mathbb{E}\left[\mathbb{P}\left(\|\nabla \boldsymbol{\rho}_{u_k}(V(u_k))\| > M/2 | W_X\right)\right] \\
&\le 2\sqrt{N \vee T}\exp(-(N \vee T)) \to 0 \qquad \text{(S.B.46)}
\end{aligned}$$

where the first inequality is by the union bound. The equality is by the law of iterated expectation. The last inequality follows from $\varepsilon = 1/\sqrt{N \vee T}$ and from Lemma S.B.5. Specifically,

[6] The sub-Gaussian norm of a sub-Gaussian random variable X is defined as $\|X\|_{\psi_2} := \inf\{c > 0 : \mathbb{E}\exp(X^2/c^2) \le 2\}$

by $C_{op} = 12CQ + 6$, $M/2 = C_{op}\sqrt{N \vee T}/2 > 3CQ\sqrt{N \vee T} \geq CQ(\sqrt{N} + \sqrt{T} + \sqrt{N \vee T})$, and the result in Lemma S.B.5 applies by letting $t = \sqrt{N \vee T}$. Note that we choose a larger C_{op} than what is needed now for later use.

Bound on $\mathcal{F}_2$. Similar to the proof of equation (S.B.1),

$$\begin{aligned}
\sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\nabla\boldsymbol{\rho}_u(V(u)) - \nabla\boldsymbol{\rho}_{u_k}(V(u_k))\| \leq & \sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\Xi_1\| + \sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\Xi_2\| \\
\leq & \varepsilon\sqrt{NT} + \sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\Xi_2\| \\
= & \sqrt{N \wedge T} + \sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\Xi_2\|
\end{aligned}$$

where Ξ_1 and Ξ_2 are defined in equations (S.B.36) and (S.B.37) in the proof of equation (S.B.1). The second inequality holds because all the elements in Ξ_1 are equal to $u - u_k$, whose magnitude is bounded by ε and the spectral norm of a matrix of all ones is equal to $\sqrt{NT}$. The last equality is by $\varepsilon = 1/\sqrt{N \vee T}$. Hence,

$$\mathbb{P}(\mathcal{F}_2 > M/2) \leq \mathbb{P}\left(\sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\Xi_2\| > M/2 - \sqrt{N \wedge T}\right). \tag{S.B.47}$$

Now we bound $\sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\Xi_2\|$. By definition, for an arbitrary matrix $N \times T$ matrix A, $\|A\| := \sup_{\|x\|_F=1} \|Ax\|_F$ where x is a $T \times 1$ vector. Suppose all the elements in A have the same sign. Then, the supremum is achieved only if all the elements in x also have the same sign and thus $\sup_{\|x\|_F=1} \|Ax\|_F = \sup_{\|x\|_F=1} \|A \cdot |x|\|_F$. Meanwhile, for a matrix B whose elements also have the same sign with $|B_{it}| \geq |A_{it}|$ for all i and t, we have $\|A \cdot |x|\|_F \leq \|B \cdot |x|\|_F$. Therefore,

$$\|A\| = \sup_{\|x\|_F=1} \|A \cdot |x|\|_F \leq \sup_{\|x\|_F=1} \|B \cdot |x|\|_F = \sup_{\|x\|_F=1} \|Bx\|_F = \|B\|. \tag{S.B.48}$$

Hence,

$$\sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_K} \|\Xi_2\| \leq \sup_{\substack{u_k-\varepsilon\leq u\leq u_k \\ u_k\in\mathcal{U}_k}} \|\Xi_2\| + \sup_{\substack{u_k\leq u\leq u_k+\varepsilon \\ u_k\in\mathcal{U}_k}} \|\Xi_2\| \leq \max_{u_k\in\mathcal{U}_k} \|\Xi_2^{(1)}\| + \max_{u_k\in\mathcal{U}_k} \|\Xi_2^{(2)}\|, \tag{S.B.49}$$

where $\Xi_2^{(1)}$ and $\Xi_2^{(2)}$ are defined in equations (S.B.40) and (S.B.41) in the proof of equation (S.B.1) and do not depend on u. To see why the second inequality holds, recall that the elements in Ξ_2 are all nonnegative when $u < u_k$ and all nonpositive when $u > u_k$, and in either case, we have $|\Xi_{2,it}| \leq |\Xi_{2,it}^{(\iota)}|$, $\iota = 1, 2$ for all i, t. Inequality (S.B.49) is thus implied by inequality (S.B.48).

Again, let us only derive the bound on $\max_{u_k\in\mathcal{U}_k}\|\Xi_2^{(1)}\|$ because the bound on $\max_{u_k\in\mathcal{U}_k}\|\Xi_2^{(2)}\|$ follows the same argument. Recall that matrix $\bar{\Xi}_2^{(1)} := (\varepsilon)_{i,t}$ is the conditional mean of $\Xi_2^{(1)}$ given W_X. Note that $\max_{u_k\in\mathcal{U}_K}\|\bar{\Xi}_2^{(1)}\| = \varepsilon\sqrt{NT} = \sqrt{N\wedge T}$ by $\varepsilon = 1/\sqrt{N\vee T}$. We have

$$\begin{aligned}
&\mathbb{P}\left(\max_{u_k\in\mathcal{U}_k}\|\Xi_2^{(1)}\| > M/4 - \sqrt{N\wedge T}/2\right)\\
\leq&\mathbb{P}\left(\max_{u_k\in\mathcal{U}_k}\|\Xi_2^{(1)} - \bar{\Xi}_2^{(1)}\| + \max_{u_k\in\mathcal{U}_k}\|\bar{\Xi}_2^{(1)}\| > C_{op}\sqrt{N\vee T}/4 - \sqrt{N\wedge T}/2\right)\\
\leq&\mathbb{P}\left(\max_{u_k\in\mathcal{U}_k}\|\Xi_2^{(1)} - \bar{\Xi}_2^{(1)}\| > C_{op}\sqrt{N\vee T}/4 - 3\sqrt{N\wedge T}/2\right)\\
\leq&K\max_{u_k\in\mathcal{U}_K}\mathbb{P}\left(\|\Xi_2^{(1)} - \bar{\Xi}_2^{(1)}\| > (C_{op}/4 - 3/2)\sqrt{N\vee T}\right)\\
\leq&\sqrt{N\vee T}\max_{u_k\in\mathcal{U}_K}\mathbb{E}\left[\mathbb{P}\left(\|\Xi_2^{(1)} - \bar{\Xi}_2^{(1)}\| > (C_{op}/4 - 3/2)\sqrt{N\vee T}\middle| W_X\right)\right]\\
\leq&\frac{2\sqrt{N\vee T}}{\exp(N\vee T)} \to 0,
\end{aligned} \tag{S.B.50}$$

where the last inequality follows the same argument for equation (S.B.46) since the entries in $(\Xi_2^{(1)}(u_k) - \bar{\Xi}_2^{(1)}(u_k))$ are independent Bernoulli with zero mean conditional on W_X. In particular, each equals $1-\varepsilon$ with probability ε and equal to $-\varepsilon$ with probability $(1-\varepsilon)$. So their sub-Gaussian norms are still bounded by Q. By $C_{op} = 12CQ+6$, $(C_{op}/4-1.5)\sqrt{N\vee T} = 3CQ\sqrt{N\vee T} \geq CQ(\sqrt{N}+\sqrt{T}+\sqrt{N\vee T})$ and thus Lemma S.B.5 applies by letting $t = \sqrt{N\vee T}$.

Similarly, we have $\mathbb{P}\left(\max_{u_k\in\mathcal{U}_k}\|\Xi_2^{(2)}\| > M/4 - \sqrt{N\wedge T}/2\right) \to 0$ as well. Combining it with equations (S.B.45), (S.B.46), (S.B.47), (S.B.49) and (S.B.50), we have the desired result. ■

Proof of Lemma S.B.2

To prove the lemma, we need the following result which helps to handle the high-dimensional Δ_L. Its proof is in this section.

Lemma S.B.6. *For all $w_1, w_2 \in \mathbb{R}$ and all $\kappa \in (0,1]$,*

$$\int_0^{w_2}\Big(\mathbb{1}(w_1\leq z) - \mathbb{1}(w_1\leq 0)\Big)dz \geq \int_0^{\kappa w_2}\Big(\mathbb{1}(w_1\leq z) - \mathbb{1}(w_1\leq 0)\Big)dz \geq 0.$$

By Knight's identity (Knight, 1998), for any two scalars w_1 and w_2,

$$\rho_u(w_1-w_2) - \rho_u(w_1) = -w_2(u - \mathbb{1}(w_1\leq 0)) + \int_0^{w_2}(\mathbb{1}(w_1\leq s) - \mathbb{1}(w_1\leq 0))ds.$$

Let $w_1 = V_{it}(u)$ and $w_2 = X_{it}'\Delta_\beta + \Delta_{L,it}$ where Δ_β and Δ_L are arbitrary fixed $p \times 1$ vector and $N \times T$ matrix, then by the definition of $V_{it}(u)$ and by the law of iterated expectation,

$$\mathbb{E}\left(-w_2(u - \mathbb{1}(w_1 \leq 0))\right) = \mathbb{E}\left[\mathbb{E}(-w_2(u - \mathbb{1}(w_1 \leq 0))|W_X)\right] = 0. \tag{S.B.51}$$

Now we lower bound the integral. By the Cauchy-Schwarz inequality, $|X_{it}'\Delta_\beta| \leq \sqrt{X_{it}'X_{it}}\|\Delta_\beta\|_F$. Define $\kappa_{it} = 1/\left(2\alpha_{NT} + \left(\sqrt{X_{it}'X_{it}}\|\Delta_\beta\|_F \vee 1\right)\right)$. By $\alpha_{NT} \geq 1$, $\kappa_{it} \in (0,1)$. By $\|\Delta_L\|_\infty \leq 2\alpha_{NT}$, we have $|\kappa_{it} \cdot (X_{it}'\Delta_\beta + \Delta_{L,it})| \leq 1$ for all i and t. Therefore, for all i and t, $\kappa_{it} \cdot (X_{it}'\Delta_\beta + \Delta_{L,it})(1 \wedge \delta) \in [-\delta, \delta]$ where $\delta > 0$ is defined in Assumption 2. Then the following holds for all i, t, and $u \in \mathcal{U}$:

$$\begin{aligned}
&\mathbb{E}\left(\int_0^{X_{it}'\Delta_\beta + \Delta_{L,it}} \Big(\mathbb{1}(V_{it}(u) \leq s) - \mathbb{1}(V_{it}(u) \leq 0)\Big) ds\right) \\
\geq &\mathbb{E}\left(\int_0^{\kappa_{it}(X_{it}'\Delta_\beta + \Delta_{L,it})(1\wedge\delta)} \Big(\mathbb{1}(V_{it}(u) \leq s) - \mathbb{1}(V_{it}(u) \leq 0)\Big) ds\right) \\
= &\mathbb{E}\left[\mathbb{E}\left(\int_0^{\kappa_{it}(X_{it}'\Delta_\beta + \Delta_{L,it})(1\wedge\delta)} \Big(\mathbb{1}(V_{it}(u) \leq s) - \mathbb{1}(V_{it}(u) \leq 0)\Big) ds \Big| W_X\right)\right] \\
= &\mathbb{E}\left[\int_0^{\kappa_{it}(X_{it}'\Delta_\beta + \Delta_{L,it})(1\wedge\delta)} \Big(F_{V_{it}(u)|W_X}(s) - F_{V_{it}(u)|W_X}(0)\Big) ds\right] \\
= &\mathbb{E}\left[\int_0^{\kappa_{it}(X_{it}'\Delta_\beta + \Delta_{L,it})(1\wedge\delta)} s f_{V_{it}(u)|W_X}(\tilde{s}(s)) ds\right] \\
\geq &\mathbb{E}\left[\frac{\kappa_{it}^2(1\wedge\delta)^2(X_{it}'\Delta_\beta + \Delta_{L,it})^2 \underline{f}}{2}\right] \\
= &\frac{(1\wedge\delta)^2 \underline{f}}{2} \cdot \mathbb{E}\left[\frac{(X_{it}'\Delta_\beta + \Delta_{L,it})^2}{\left(2\alpha_{NT} + \left(\sqrt{X_{it}'X_{it}}\|\Delta_\beta\|_F \vee 1\right)\right)^2}\right],
\end{aligned} \tag{S.B.52}$$

where $\underline{f}$ is defined in Assumption 2. The first inequality is by Lemma S.B.6. The equality immediately after it is by the law of iterated expectation. The third equality is by the mean value theorem where $\tilde{s}(s)$ is a mean value. The inequality following is by Assumption 2 and by $\kappa_{it} \cdot (X_{it}'\Delta_\beta + \Delta_{L,it})(1 \wedge \delta) \in [-\delta, \delta]$. The last equality is by the definition of κ_{it}. Now we lower bound the expectation on the right hand side of the last equality. We have

$$\begin{aligned}
&\mathbb{E}\left[\frac{(X_{it}'\Delta_\beta + \Delta_{L,it})^2}{\left(2\alpha_{NT} + \left(\sqrt{X_{it}'X_{it}}\|\Delta_\beta\|_F \vee 1\right)\right)^2}\right] \\
\geq &\mathbb{E}\left[\frac{\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)(X_{it}'\Delta_\beta + \Delta_{L,it})^2}{\left(2\alpha_{NT} + \left(\sqrt{X_{it}'X_{it}}\|\Delta_\beta\|_F \vee 1\right)\right)^2}\right]
\end{aligned}$$

$$
\begin{aligned}
\geq & \frac{\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)(X_{it}'\Delta_\beta + \Delta_{L,it})^2\right]}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_\beta\|_F/\gamma \vee 1))^2} \\
= & \frac{\mathbb{E}\left[(X_{it}'\Delta_\beta + \Delta_{L,it})^2\right]}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_\beta\|_F/\gamma \vee 1))^2} - \frac{\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)\cdot(X_{it}'\Delta_\beta + \Delta_{L,it})^2\right]}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_\beta\|_F/\gamma \vee 1))^2} \\
\geq & \frac{\mathbb{E}\left[(X_{it}'\Delta_\beta + \Delta_{L,it})^2\right]}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_\beta\|_F/\gamma \vee 1))^2} \\
& - 2 \times \frac{\|\Delta_\beta\|_F^2 \mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)\cdot X_{it}'X_{it}\right] + \Delta_{L,it}^2 \mathbb{P}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_\beta\|_F/\gamma \vee 1))^2} \\
\geq & \frac{\mathbb{E}\left[(X_{it}'\Delta_\beta + \Delta_{L,it})^2\right] - \varepsilon_{NT}\left(\|\Delta_\beta\|_F^2 + \Delta_{L,it}^2\right)}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_\beta\|_F/\gamma \vee 1))^2} \qquad \text{(S.B.53)}
\end{aligned}
$$

where the third inequality is by Cauchy-Schwarz. The last inequality is by substituting $\varepsilon_{NT} := 2\left(\max_{i,t}\mathbb{E}\left(\mathbb{1}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)\cdot X_{it}'X_{it}\right) \vee \max_{i,t}\mathbb{P}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)\right)$ into the right hand side. We obtain the desired results by combining equations (S.B.51), (S.B.52) and (S.B.53) and summing them over i and t.

Now we show that $\varepsilon_{NT} = 0$ for large enough N and T under Assumption 3 i) and $\varepsilon_{NT} \to 0$ under Assumption 3 ii). Since $p = o((N \wedge T)/(\log(NT)\alpha_{NT}^2)$ under both Assumption 3 i) and ii), $\gamma = O(\alpha_{NT}^2\sqrt{\log(NT)}/\sqrt{N \wedge T})$ because $C_\lambda = O\left(\sqrt{\log(NT)}\right)$. Therefore, $\sqrt{p} = o(\alpha_{NT}/\gamma)$.

Assumption 3 i). When $\max_{j=1,\dots,p;i,t}|X_{j,it}| \leq \sqrt{C_X}$, $\sqrt{X_{it}'X_{it}} \leq \sqrt{C_X p}$. So $\mathbb{1}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right) \leq \mathbb{1}\left(\sqrt{C_X p} > \alpha_{NT}/\gamma\right) = 0$ for all i and t for large enough N and T because $\sqrt{p} = o(\alpha_{NT}/\gamma)$. Therefore, $\max_{i,t}\mathbb{E}\left(\mathbb{1}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)\cdot X_{it}'X_{it}\right)$ and $\max_{i,t}\mathbb{P}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)$ are both equal to 0 for large enough N and T.

Assumption 3 ii). We first show that $\max_{i,t}\mathbb{P}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right) \to 0$:

$$
\begin{aligned}
\max_{i,t}\mathbb{P}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right) \leq & p \max_{j=1,\dots,p;i,t}\mathbb{P}\left(|X_{j,it}| > \frac{\alpha_{NT}}{\sqrt{p+1}\gamma}\right) \\
\leq & \frac{p\max_{j=1,\dots,p;i,t}E(\varphi(|X_{j,it}|))}{\varphi\left(\frac{\alpha_{NT}}{\sqrt{p+1}\gamma}\right)} \to 0
\end{aligned}
$$

where the first inequality is by the union bound. The second inequality is by the Markov's inequality. Since the order of $\alpha_{NT}/(\sqrt{p+1}\gamma)$ is at least $(\sqrt{N \wedge T}/(\sqrt{(p+1)\log(NT)}\alpha_{NT}))$, convergence is by $p/\varphi\left(\sqrt{N \wedge T}/(\sqrt{(p+1)\log(NT)}\alpha_{NT})\right) \to 0$ and $\max_{j=1,\dots,p;i,t}E(\varphi(|X_{j,it}|)) < C_\varphi$ by Assumption 3 ii).

Now we show that $\max_{i,t}\mathbb{E}\left(\mathbb{1}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)\cdot X_{it}'X_{it}\right)$ converges to zero as well.

We have

$$
\begin{aligned}
&\max_{i,t} \mathbb{E}\left(\mathbb{1}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right) \cdot X_{it}'X_{it}\right) \\
\leq &\max_{i,t} \left[\left(\mathbb{E}[(X_{it}'X_{it})^{1+\eta/2}]\right)^{2/(2+\eta)} \cdot \left(\mathbb{P}\left(\sqrt{X_{it}'X_{it}} \geq \alpha_{NT}/\gamma\right)\right)^{\eta/(2+\eta)}\right] \\
\leq &p \max_{j=1,\ldots,p;i,t} \left(\mathbb{E}\left(|X_{j,it}|^{2+\eta}\right)\right)^{2/(2+\eta)} \max_{i,t} \left[\mathbb{P}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)\right]^{\eta/(2+\eta)} \\
\leq &p \left(\max_{j=1,\ldots,p;i,t} \mathbb{E}\left(|X_{j,it}|^{2+\eta}\right)\right)^{2/(2+\eta)} \left(\frac{p \max_{j=1,\ldots,p;i,t} E(\varphi(|X_{j,it}|))}{\varphi\left(\frac{\alpha_{NT}}{\sqrt{p+1}\gamma}\right)}\right)^{\eta/(2+\eta)} \\
\leq &\left(\max_{j=1,\ldots,p;i,t} E(\varphi(|X_{j,it}|))\right) p \cdot \left(\frac{p}{\varphi\left(\frac{\alpha_{NT}}{\sqrt{p+1}\gamma}\right)}\right)^{\eta/(2+\eta)} \to 0
\end{aligned}
$$

where the first inequality is by Hölder's. The second is by Minkowski. The third inequality follows the same derivation of the convergence of $\max_{i,t} \mathbb{P}\left(\sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma\right)$. The last inequality is by $\varphi(x) \geq x^{2+\eta}$ for all $x \geq 0$. Convergence is by Assumption 3 ii) and by the order of γ. ■

Proof of Lemma S.B.3

Let $A(u) \coloneqq C_{\Phi X}\kappa_3 \|\Delta_\beta\|_F - \varepsilon_0 M(u)$ where ε_0 is the same as in Assumption 4 i). We have

$$
\begin{aligned}
&\inf_{u\in\mathcal{U}} \left[M(u)^\perp - C_{\Phi X}\sqrt{3r(u)}\kappa_1\left(M(u) + \kappa_3 \|\Delta_\beta\|_F\right)\right] \\
= \min &\left\{\inf_{u\in\{u\in\mathcal{U}:A(u)<0\}} \left[M(u)^\perp - C_{\Phi X}\sqrt{3r(u)}\kappa_1\left(M(u) + \kappa_3 \|\Delta_\beta\|_F\right)\right],\right. \\
&\left.\inf_{u\in\{u\in\mathcal{U}:A(u)\geq 0\}} \left[M(u)^\perp - C_{\Phi X}\sqrt{3r(u)}\kappa_1\left(M(u) + \kappa_3 \|\Delta_\beta\|_F\right)\right]\right\}.
\end{aligned} \tag{S.B.54}
$$

First consider the case when $A(u) < 0$. The following inequalities hold:

$$
\begin{aligned}
&\inf_{u\in\{u\in\mathcal{U}:A(u)<0\}} \left[M(u)^\perp - C_{\Phi X}\sqrt{3r(u)}\kappa_1\left(M(u) + \kappa_3 \|\Delta_\beta\|_F\right)\right] \\
\geq &\inf_{u\in\{u\in\mathcal{U}:A(u)<0\}} \left(M(u)^\perp - (C_{\Phi X} + \varepsilon_0)\sqrt{3r(u)}\kappa_1 M(u)\right) \\
\geq &\inf_{u\in\mathcal{U}} \left(M(u)^\perp - (C_{\Phi X} + \varepsilon_0)\sqrt{3r(u)}\kappa_1 M(u)\right) \\
\geq &0,
\end{aligned} \tag{S.B.55}
$$

where the last inequality is by Assumption 4 i).

Now let us consider the case when $A(u) \geq 0$. We have

$$\left(M(u)\right)^2 + \left(M(u)^{\perp}\right)^2 = \mathbb{E}\left\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\right\|_F^2 \geq NT\sigma_{min}^2 \|\Delta_\beta\|_F^2 \tag{S.B.56}$$

where the inequality is by equation (S.B.10). Meanwhile, recall that the order of κ_3 is the same as κ_2 which is $o(N \wedge T)$ uniformly in u. So $M(u) = o(N \wedge T)\|\Delta_\beta\|_F$ by $A \geq 0$. So, by equation (S.B.56), the order of $M(u)^{\perp}$ must be at least $\sqrt{NT}\|\Delta_\beta\|_F$ uniformly in $u \in \{u \in \mathcal{U} : A(u) \geq 0\}$, dominating $\kappa_3\|\Delta_\beta\|_F$. Therefore, there must exist N_3 and T_3 which do not depend on u such that for all $N > T_3$ and $T > T_3$,

$$\begin{aligned}
&\inf_{u \in \{u \in \mathcal{U}: A(u) \geq 0\}} \left[M(u)^{\perp} - C_{\Phi X}\sqrt{3r(u)}\kappa_1\left(M(u) + \kappa_3 \left\|\Delta_\beta\right\|_F\right)\right] \\
\geq &\inf_{u \in \{u \in \mathcal{U}: A(u) \geq 0\}} \left[M(u)^{\perp} - \kappa_1\sqrt{3r(u)}C_{\Phi X}\left(\frac{C_{\Phi X}}{\varepsilon_0} + 1\right)\kappa_3 \left\|\Delta_\beta\right\|_F\right] \\
\geq &0.
\end{aligned} \tag{S.B.57}$$

Combining equations (S.B.54), (S.B.55) and (S.B.57), we have the desired result. ■

Proof of Lemma S.B.4

The main argument of the proof follows the proof of Lemma 5 in Belloni and Chernozhukov (2011). The major difference is that we need to handle the matrix component Δ_L.

Let

$$\mathcal{A}(\gamma) := \sup_{\substack{u \in \mathcal{U} \\ (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2}} \left|\mathbb{G}_u\left(\rho_u\left(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\right) - \rho_u(V_{it}(u))\right)\right|.$$

For arbitrary random variables Z_{it}s and a function f, denote the symmetrized version of $\mathbb{G}(f(Z_{it}))$ by $\mathbb{G}^0(f(Z_{it})) := (\sum_{i,t} f(Z_{it})\varepsilon_{it})/\sqrt{NT}$ where $(\varepsilon_{it})_{i,t}$ is a Rademacher sequence independent of $(\{V(u)\}_{u \in (0,1)}, W)$.

Similar to Belloni and Chernozhukov (2011) and Chao et al. (2021), for any fixed Δ_β and Δ_L with $\|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2$ and any $u \in \mathcal{U}$, we have the following bound on the conditional variance of the process by noting that the check function is a contraction:

$$\mathrm{Var}\left(\mathbb{G}_u\left(\rho_u\left(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\right) - \rho_u(V_{it}(u))\right)\right)$$

$$
\begin{aligned}
&\leq \frac{1}{NT}\mathbb{E}\left(\left\|\sum_{j=1}^{p} X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2\right) \\
&\leq \frac{2p}{NT}\sum_{i,t}\max_{j=1,\ldots,p}\mathbb{E}(X_{j,it}^2)\|\Delta_\beta\|_F^2 + \frac{2}{NT}\|\Delta_L\|_F^2 \\
&\leq 2p\left(\max_{i,t,j=1,\ldots,p}\mathbb{E}(X_{j,it}^2)\right)\|\Delta_\beta\|_F^2 + \frac{2}{NT}\|\Delta_L\|_F^2 \\
&\leq 2\left(p\max_{i,t,j=1,\ldots,p}\mathbb{E}(X_{j,it}^2)\vee 1\right)\left(\|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2\right) \\
&\leq 2\left(p\max_{i,t,j=1,\ldots,p}\mathbb{E}(\varphi(|X_{j,it}|))\vee 1\right)\left(\|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2\right) \\
&\leq 2\left(pC_\varphi \vee 1\right)\gamma^2, \qquad \text{(S.B.58)}
\end{aligned}
$$

where $\varphi(\cdot)$ and C_φ are defined in Assumption 3. Let

$$
s = C_{sup}\sqrt{\log(NT)}\left(\sqrt{p\log((p+1)NT)} \vee \sqrt{\bar{r}(N\vee T)}\right)\gamma.
$$

Since $\mathbb{E}(\mathbb{G}_u\left(\rho_u\left(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\right) - \rho_u(V_{it}(u))\right)) = 0$ by construction, with inequality (S.B.58) we can apply the symmetrization lemma for probability, for instance Lemma 2.3.7 in van der Vaart and Wellner (1996):

$$
\mathbb{P}\left(\mathcal{A}(\gamma) > s\right) \leq \frac{2\mathbb{P}\left(\mathcal{A}^0(\gamma) > \frac{s}{4}\right)}{1 - 8\left(pC_\varphi \vee 1\right)\gamma^2/s^2} \leq 4\mathbb{P}\left(\mathcal{A}^0(\gamma) > \frac{s}{4}\right) \qquad \text{(S.B.59)}
$$

where $\mathcal{A}^0(\gamma)$ is the symmetrized version of $\mathcal{A}(\gamma)$ by replacing $\mathbb{G}_u$ with its symmetrized version $\mathbb{G}^0$. The first inequality is by Lemma 2.3.7 in van der Vaart and Wellner (1996) and Chebyshev's inequality, and by the bound on the conditional variance (S.B.58). The second inequality holds because $\gamma\sqrt{(pC_\varphi \vee 1)}/s \to 0$ by the definition of s for any fixed C_{sup}.

Let Ω_1 be the event that $\max_{1\leq j\leq p}\|X_j\|_F^2 \leq C_X NT$ where C_X is as in Assumption 1. We have

$$
\mathbb{P}\left(\mathcal{A}^0(\gamma) > \frac{s}{4}\right) \leq \mathbb{P}\left(\mathcal{A}^0(\gamma) > \frac{s}{4}\middle|\Omega_1\right) + \mathbb{P}(\Omega_1^c). \qquad \text{(S.B.60)}
$$

Since $\mathbb{P}(\Omega_1^c) \to 0$ under Assumption 1, we only need to show that $\mathbb{P}(\mathcal{A}^0(\gamma) > s/4|\Omega_1) \to 0$ under our choice of s for some C_{sup}.

Consider the random variable $\rho_u\left(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\right) - \rho_u(V_{it}(u))$:

$$
\rho_u\left(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\right) - \rho_u(V_{it}(u)) = -u\cdot\left(X_{it}'\Delta_\beta + \Delta_{L,it}\right) + \delta_{it}\left(X_{it}'\Delta_\beta + \Delta_{L,it}, u\right)
$$

where $\delta_{it}\left(X_{it}'\Delta_\beta + \Delta_{L,it}, u\right) = \left(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\right)_- - \left(V_{it}(u)\right)_-$. Let

$$\mathcal{B}_1^0(\gamma) \coloneqq \sup_{\|\Delta_\beta\|_F^2 \le \gamma^2} \left|\mathbb{G}^0\left(X_{it}'\Delta_\beta\right)\right|,$$

$$\mathcal{B}_2^0(\gamma) \coloneqq \sup_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \le \gamma^2}} \left|\mathbb{G}^0\left(\Delta_{L,it}\right)\right|,$$

and

$$\mathcal{C}^0(\gamma) \coloneqq \sup_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \le \gamma^2}} \left|\mathbb{G}^0\left(\delta_{it}\left(X_{it}'\Delta_\beta + \Delta_{L,it}, u\right)\right)\right|,$$

then $\mathcal{A}^0(\gamma) \le \mathcal{B}_1^0(\gamma) + \mathcal{B}_2^0(\gamma) + \mathcal{C}^0(\gamma)$ with probability one. Hence,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{A}^0(\gamma) \ge \frac{s}{4}\Big|\Omega_1\right) \\
\le&\mathbb{P}\left(\mathcal{B}_1^0(\gamma) + \mathcal{B}_2^0(\gamma) + \mathcal{C}^0(\gamma) \ge \frac{s}{4}\Big|\Omega_1\right) \\
\le&3\max\left\{\mathbb{P}\left(\mathcal{B}_1^0(\gamma) \ge \frac{s}{12}\Big|\Omega_1\right), \mathbb{P}\left(\mathcal{B}_2^0(\gamma) \ge \frac{s}{12}\Big|\Omega_1\right), \mathbb{P}\left(\mathcal{C}^0(\gamma) \ge \frac{s}{12}\Big|\Omega_1\right)\right\}.
\end{aligned} \tag{S.B.61}$$

We now derive upper bounds on $\mathcal{B}_1^0(\gamma)$, $\mathcal{B}_2^0(\gamma)$ and $\mathcal{C}^0(\gamma)$ respectively.

Bound on $\mathcal{B}_1^0(\gamma)$. The derivation of the bound on $\mathcal{B}_1^0(\gamma)$ follows Belloni and Chernozhukov (2011) closely. We present the proof here for completeness. For some $K_1 > 0$, by Markov's inequality,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{B}_1^0(\gamma) > K_1\Big|W_X, \Omega_1\right) \\
\le&\min_{\tau\ge 0} e^{-\tau K_1}\mathbb{E}\left[\exp\left(\tau\mathcal{B}_1^0(\gamma)\right)\Big|W_X, \Omega_1\right] \\
\le&\min_{\tau\ge 0} e^{-\tau K_1}\mathbb{E}\left[\exp\left(\tau \sup_{\|\Delta_\beta\|_F^2\le\gamma^2} \|\Delta_\beta\|_1 \cdot \max_{1\le j\le p}|\mathbb{G}^0(X_{j,it})|\right)\Bigg|W_X, \Omega_1\right] \\
\le&2p\min_{\tau\ge 0} e^{-\tau K_1}\max_{1\le j\le p}\mathbb{E}\left[\exp\left(\tau\sqrt{p}\gamma\cdot\mathbb{G}^0(X_{j,it})\right)\Big|W_X, \Omega_1\right] \\
\le&2p\min_{\tau\ge 0} e^{-\tau K_1}\exp\left(\frac{\tau^2 p\gamma^2 C_X}{2}\right)
\end{aligned}$$

where the third inequality follows from the fact that $\mathbb{E}[\max_{1\le j\le p}\exp(|z_j|)] \le 2p\max_{1\le j\le p}\mathbb{E}[\exp(z_j)]$ for a symmetric random variable z_j (Belloni and Chernozhukov, 2011). The last inequality is by an intermediate step in the proof of Hoeffding's inequality (e.g. van der Vaart and Wellner (1996) p.100) and by Ω_1. Hence, by setting $\tau = K_1/(p\gamma^2 C_X)$ and $K_1 = \sqrt{2pC_X\log((p+1)NT)}\cdot\gamma$ (if $p = 0$, then let τ be any constant and the bound is equal to

0), we have

$$\mathbb{P}\left(\mathcal{B}_1^0(\gamma) > K_1 \Big| W_X, \Omega_1\right) \leq 2p \exp\left(-\frac{K_1^2}{2p\gamma^2 C_X}\right) = \frac{2p}{(p+1)NT} \to 0.$$

Therefore,

$$\mathbb{P}\left(\mathcal{B}_1^0 > K_1 \Big| \Omega_1\right) = \mathbb{E}\left[\mathbb{P}\left(\mathcal{B}_1^0 > K_1 | W_X, \Omega_1\right) \Big| \Omega_1\right] \leq \mathbb{E}\left(\frac{2}{NT}\right) \to 0. \tag{S.B.62}$$

Bound on $\mathcal{B}_2^0(\gamma)$. Recall that $\{\varepsilon_{it}\}_{i,t}$ is the Rademacher sequence in the symmetrized process. Let $\boldsymbol{\varepsilon}$ be the $N \times T$ matrix $(\varepsilon_{it})_{i,t}$. Then with probability one, by defining $C_{Cone} := 5\sqrt{2C_X}/(C_\lambda C_{op})$, we have

$$\begin{aligned}
&\mathcal{B}_2^0(\gamma) \\
=&\frac{1}{\sqrt{NT}} \sup_{\substack{u \in \mathcal{U} \\ (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2}} |\sum_{i,t} \varepsilon_{it} \Delta_{L,it}| \\
=&\frac{1}{\sqrt{NT}} \sup_{\substack{u \in \mathcal{U} \\ (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2}} |\langle \boldsymbol{\varepsilon}, \Delta_L \rangle| \\
\leq&\frac{1}{\sqrt{NT}} \|\boldsymbol{\varepsilon}\| \cdot \sup_{\substack{u \in \mathcal{U} \\ (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2}} \|\Delta_L\|_* \\
\leq&\frac{1}{\sqrt{NT}} \|\boldsymbol{\varepsilon}\| \cdot \sup_{\|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2} \left(C_{Cone}\sqrt{p(N \wedge T)\log((p+1)NT)} \|\Delta_\beta\|_F + \frac{2(1+C_\lambda)}{C_\lambda} \|\mathcal{P}_{\Phi(u)} \Delta_L\|_*\right) \\
\leq&\frac{1}{\sqrt{NT}} \|\boldsymbol{\varepsilon}\| \cdot \sup_{\|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2} \left(C_{Cone}\sqrt{p(N \wedge T)\log((p+1)NT)} \|\Delta_\beta\|_F + \frac{2\sqrt{3\bar{r}}(1+C_\lambda)}{C_\lambda} \|\Delta_L\|_F\right) \\
\leq&\frac{2}{\sqrt{NT}} \|\boldsymbol{\varepsilon}\| \cdot \left(\left(C_{Cone}\sqrt{p(N \wedge T)\log((p+1)NT)}\right) \vee \left(\frac{2(1+C_\lambda)}{C_\lambda}\sqrt{3\bar{r}NT}\right)\right)\gamma,
\end{aligned}$$

where the second inequality is by the definition of cone $\mathcal{R}_u$ and the third inequality is by equation (3.9). Since $\boldsymbol{\varepsilon}$ has i.i.d. mean 0 Bernoulli entries, there exists a constant $C_{Sp} > 1$ such that for $\mathbb{P}\left(\|\boldsymbol{\varepsilon}\| > C_{Sp}\sqrt{N \vee T}/2\right) \to 0$ by Lemma S.B.5. Let

$$K_2 := C_{Sp}\left(\left(C_{Cone}\sqrt{p\log((p+1)NT)}\right) \vee \left(2(1+C_\lambda)\sqrt{N \vee T}\sqrt{3\bar{r}}/C_\lambda\right)\right)\gamma,$$

we have

$$\mathbb{P}\left(\mathcal{B}_2^0(\gamma) > K_2 \middle| \Omega_1\right) \leq \mathbb{P}\left(\|\boldsymbol{\varepsilon}\| > C_{Sp}\sqrt{N \vee T}/2\right) \to 0. \tag{S.B.63}$$

Bound on $\mathcal{C}^0(\gamma)$. By $\gamma = o(1)$, it is smaller than one for sufficiently large N and T. Then let $\mathcal{U}_l = \{u_1, ..., u_l\}$ be an ϵ-net of $\mathcal{U}$ where $\epsilon = \gamma/\sqrt{p+1}$ and $\epsilon l \leq 1$. For any $\bar{u} \in \mathcal{U}$, we have the identity

$$\begin{aligned}\delta_{it}(X_{it}'\Delta_\beta + \Delta_{L,it}, u) =& \delta_{it}[X_{it}'(\Delta_\beta + \beta_0(u) - \beta_0(\bar{u})) + \Delta_{L,it} + L_{0,it}(u) - L_{0,it}(\bar{u}), \bar{u}] \\ &- \delta_{it}[X_{it}'(\beta_0(u) - \beta_0(\bar{u})) + L_{0,it}(u) - L_{0,it}(\bar{u}), \bar{u}].\end{aligned}$$

Then by the triangle inequality, with probability one we have

$$\begin{aligned}\mathcal{C}^0(\gamma) \leq & \sup_{\substack{u \in \mathcal{U}, |u-\bar{u}|<\epsilon, \bar{u} \in \mathcal{U}_l \\ (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2}} \left|\mathbb{G}^0\left(\delta_{it}[X_{it}'(\Delta_\beta + \beta_0(u) - \beta_0(\bar{u})) + \Delta_{L,it} + L_{0,it}(u) - L_{0,it}(\bar{u}), \bar{u}]\right)\right| \\ & + \sup_{\substack{u \in \mathcal{U}, |u-\bar{u}|<\epsilon, \bar{u} \in \mathcal{U}_l \\ (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2}} \left|\mathbb{G}^0\left(\delta_{it}[X_{it}'(\beta_0(u) - \beta_0(\bar{u})) + L_{0,it}(u) - L_{0,it}(\bar{u}), \bar{u}]\right)\right|. \end{aligned} \tag{S.B.64}$$

We will proceed by treating $\Delta_\beta + \beta_0(u) - \beta_0(\bar{u})$ and $\beta_0(u) - \beta_0(\bar{u})$ as new Δ_βs, and $\Delta_L + L_0(u) - L_0(\bar{u})$ and $L_0(u) - L_0(\bar{u})$ as new Δ_Ls. However, they may no longer lie in the ball $\|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq \gamma^2$ and in cone $\mathcal{R}_u$. So, we need to first expand these two sets.

We first expand the ball. For Δ_β, by Assumption 5, by $\epsilon = \gamma/\sqrt{p+1}$ and by $\|\Delta_\beta\|_F \leq \gamma$, we have $\|\Delta_\beta + \beta_0(u) - \beta_0(\bar{u})\|_F^2 \leq 2(1 + \zeta_X^2/(p+1))\gamma^2$ and $\|\beta_0(u) - \beta_0(\bar{u})\|_F^2 \leq \zeta_X^2\gamma^2/(p+1)$ for all $|u - \bar{u}| \leq \epsilon$. Similarly, for $\|\Delta_L\|_F \leq \sqrt{NT}\gamma$, under equation (3.2), $\|\Delta_L + L_0(u) - L_0(\bar{u})\|_F^2/NT \leq 2(1 + \zeta_L^2/(p+1))\gamma^2$ while $\|L_0(u) - L_0(\bar{u})\|_F^2/NT \leq \zeta_L^2\gamma^2/(p+1)$ for all $|u - \bar{u}| \leq \epsilon$. Therefore, we need to expand the ball to be $\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2/NT \leq 2(1 + (\zeta_X^2 + \zeta_L^2)/(p+1))\gamma^2$. For simplicity, let $C_\zeta \coloneqq 1 + (\zeta_X^2 + \zeta_L^2)/(1+p)$. Since $\zeta_X = O(\sqrt{p})$, $C_\zeta = O(1)$.

Next, let us expand $\mathcal{R}_u$. Since $\text{rank}(L_0(u) - L_0(\bar{u})) \leq r(u) + r(\bar{u}) \leq 2\bar{r}$, we have $\|L_0(u) - L_0(\bar{u})\|_* \leq \sqrt{2\bar{r}}\|L_0(u) - L_0(\bar{u})\|_F \leq \sqrt{2\bar{r}}\zeta_L\sqrt{NT}\gamma/\sqrt{p+1}$ by $\epsilon = \gamma/\sqrt{p+1}$ for all $|u - \bar{u}| \leq \epsilon$. Similarly, for $(\Delta_\beta, \Delta_L) \in \mathcal{R}_u$ and $\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2/NT \leq 2C_\zeta\gamma^2$, the following holds for all $u \in \mathcal{U}$ and all $|u - \bar{u}| \leq \epsilon$ with probability one,

$$\begin{aligned}&\|\Delta_L + L_0(u) - L_0(\bar{u})\|_* \\ \leq& \|\Delta_L\|_* + \|L_0(u) - L_0(\bar{u})\|_* \\ \leq& \frac{2(1 + C_\lambda)}{C_\lambda}\|\mathcal{P}_{\Phi(u)}\Delta_L(u)\|_* + C_{Cone}\sqrt{p\log((p+1)NT)(N \wedge T)}\|\Delta_\beta\|_F + \sqrt{2\bar{r}}\zeta_L\sqrt{NT}\gamma/\sqrt{p+1}\end{aligned}$$

$$\leq \left(\frac{2(1+C_\lambda)\sqrt{6C_\zeta}}{C_\lambda} + \sqrt{2}\zeta_L \right) \sqrt{NT\bar{r}}\gamma + C_{Cone}\sqrt{p\log((p+1)NT)(N \wedge T)}\sqrt{2C_\zeta}\gamma$$

where the second inequality follows from the definition of $\mathcal{R}_u$. The last inequality is by $\gamma/\sqrt{p+1} \leq \gamma$. Let

$$\bar{\mathcal{R}} = \Bigg\{ \Delta_L \in \mathbb{R}^{N\times T} : \\ \|\Delta_L\|_* \leq \left(\frac{2(1+C_\lambda)\sqrt{6C_\zeta}}{C_\lambda} + \sqrt{2}\zeta_L \right) \sqrt{NT\bar{r}}\gamma + C_{Cone}\sqrt{2C_\zeta p\log((p+1)NT)(N \wedge T)}\gamma \Bigg\}.$$

Therefore, in the intersection of the ball $\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2/NT \leq 2C_\zeta\gamma^2$ and $\mathcal{R}_u$ for all $u \in \mathcal{U}$, the matrices Δ_L, $(\Delta_L + L_0(u) - L_0(\bar{u}))$ and $(L_0(u) - L_0(\bar{u}))$ are all in $\bar{\mathcal{R}}$ for all $u \in \mathcal{U}$ and $|u - \bar{u}| \leq \epsilon$. Hence, inequality (S.B.64) implies that,

$$\mathcal{C}^0(\gamma) \leq 2 \cdot \sup_{\substack{\bar{u}\in\mathcal{U}_l, \Delta_L \in \bar{\mathcal{R}} \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq 2C_\zeta\gamma^2}} \left| \mathbb{G}^0\Big(\delta_{it}(X_{it}'\Delta_\beta + \Delta_{L,it}, \bar{u})\Big) \right| := 2\mathcal{C}^1(\gamma). \tag{S.B.65}$$

Define the following event:

$$\Omega_2 := \Bigg\{ \sup_{\Delta_L \in \bar{\mathcal{R}}} \left| \mathbb{G}^0\left(\Delta_{L,it}\right) \right| \\ \leq C_{Sp}\left[\left(C_{Cone}\sqrt{2C_\zeta p\log((p+1)NT)} \right) \vee \left(\sqrt{N \vee T} \left(\frac{\sqrt{6C_\zeta}2(1+C_\lambda)}{C_\lambda} + \sqrt{2}\zeta_L \right) \sqrt{\bar{r}} \right) \right] \gamma \Bigg\}.$$

Now we can derive the upper bound on $\mathcal{C}^0(\gamma)$. For some $K_3 > 0$, by equation (S.B.65),

$$\mathbb{P}\left(\mathcal{C}^0(\gamma) \geq 2K_3 \Big| W_X, \Omega_1\right) \leq \mathbb{P}\left(\mathcal{C}^1(\gamma) \geq K_3 \Big| W_X, \Omega_1\right) \leq e^{-\tau' K_3}\mathbb{E}\left[e^{\tau'\mathcal{C}^1(\gamma)} \Big| W_X, \Omega_1 \right], \tag{S.B.66}$$

where the last equality is by Markov's inequality for some $\tau' > 0$. For $\mathbb{E}\left[e^{\tau'\mathcal{C}^1(\gamma)} \Big| W_X, \Omega_1 \right]$, by $l \leq 1/\epsilon = \sqrt{p+1}/\gamma$, we have

$$\mathbb{E}\left[e^{\tau'\mathcal{C}^1(\gamma)} | W_X, \Omega_1 \right] \\ \leq \frac{\sqrt{p+1}}{\gamma} \max_{\bar{u}\in\mathcal{U}_l} \mathbb{E}\left[\exp\left(\tau' \sup_{\substack{\Delta_L \in \bar{\mathcal{R}} \\ \|\Delta_\beta\|_F^2 + \frac{1}{NT}\|\Delta_L\|_F^2 \leq 2C_\zeta\gamma^2}} \left| \mathbb{G}^0\left(\delta_{it}(X_{it}'\Delta_\beta + \Delta_{L,it}, \bar{u})\right) \right| \right) \Bigg| W_X, \Omega_1 \right]$$

$$
\begin{aligned}
&\le \frac{\sqrt{p+1}}{\gamma}\mathbb{E}\left[\exp\left(2\tau' \sup_{\substack{\Delta_L\in\bar{\mathcal{R}}\\ \|\Delta_\beta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2\le 2C_\zeta\gamma^2}} \left|\mathbb{G}^0\left(X_{it}'\Delta_\beta+\Delta_{L,it}\right)\right|\right)\middle| W_X,\Omega_1\right]\\
&\le \frac{\sqrt{p+1}}{\gamma}\mathbb{E}\left[\exp\left(2\tau' \sup_{\|\Delta_\beta\|_F^2\le 2C_\zeta\gamma^2}\left|\mathbb{G}^0(X_{it}'\Delta_\beta)\right| + 2\tau' \sup_{\substack{\Delta_L\in\bar{\mathcal{R}}\\ \|\Delta\|_F^2+\frac{1}{NT}\|\Delta_L\|_F^2\le 2C_\zeta\gamma^2}}\left|\mathbb{G}^0(\Delta_{L,it})\right|\right)\middle| W_X,\Omega_1\right]\\
&\le \frac{\sqrt{p+1}}{\gamma}\left(\mathbb{E}\left[\exp\left(4\tau' \sup_{\|\Delta_\beta\|_F^2\le 2C_\zeta\gamma^2}\left|\mathbb{G}^0(X_{it}'\Delta_\beta)\right|\right)\middle| W_X,\Omega_1\right]\right)^{\frac{1}{2}} \times \left(\mathbb{E}\left[\exp\left(4\tau'\sup_{\Delta_L\in\bar{\mathcal{R}}}\left|\mathbb{G}^0(\Delta_{L,it})\right|\right)\right]\right)^{\frac{1}{2}}\\
&\le \frac{\sqrt{2}(p+1)}{\gamma}\exp\left(8\tau'^2 pC_\zeta\gamma^2C_X\right)\left(\mathbb{E}\left[\exp\left(4\tau'\sup_{\Delta_L\in\bar{\mathcal{R}}}\left|\mathbb{G}^0(\Delta_{L,it})\right|\right)\right]\right)^{\frac{1}{2}}, \qquad \text{(S.B.67)}
\end{aligned}
$$

where the second inequality is by Theorem 4.12 of Ledoux and Talagrand (1991) and by contractivity of $\delta_{it}(\cdot)$ with $\delta_{it}(0)=0$. The fourth inequality is by Cauchy-Schwarz. The last inequality follows the same steps as the derivation of the bound on $\mathcal{B}_1^0(\gamma)$.

By $\mathbb{G}^0(\Delta_{L,it}) \le \|\boldsymbol{\varepsilon}\| \cdot \|\Delta_L\|_*/\sqrt{NT}$ and by the definition of $\bar{\mathcal{R}}$, Ω_2 does not hold only if $\|\boldsymbol{\varepsilon}\| > C_{sp}(N\vee T)/2$. Hence, by Lemma S.B.5, $\mathbb{P}(\Omega_2^c) \le 2\exp(-(N\vee T))$. Note that $\mathbb{G}^0(\Delta_{L,it}) \le \|\Delta_L\|_*$ almost surely since $\|\boldsymbol{\varepsilon}\| \le \|\boldsymbol{\varepsilon}\|_F = \sqrt{NT}$. Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(4\tau'\sup_{\Delta_L\in\bar{\mathcal{R}}}\left|\mathbb{G}^0(\Delta_{L,it})\right|\right)\right]\\
\le&\mathbb{P}(\Omega_2)\exp\left(4\tau' C_{Sp}\left[\left(C_{Cone}\sqrt{2C_\zeta p\log((p+1)NT)}\right)\vee\left(\sqrt{N\vee T}\left(\frac{\sqrt{6C_\zeta}2(1+C_\lambda)}{C_\lambda}+\sqrt{2}\zeta_L\right)\sqrt{\bar{r}}\right)\right]\gamma\right)\\
&+2\exp(-(N\vee T))\exp\left(4\tau'\left[C_{Cone}\sqrt{2C_\zeta p\log((p+1)NT)(N\wedge T)}+\left(\frac{\sqrt{6C_\zeta}2(1+C_\lambda)}{C_\lambda}+\sqrt{2}\zeta_L\right)\sqrt{NT\bar{r}}\right]\gamma\right).
\end{aligned}
$$

Let

$$
\tau' = \frac{\sqrt{\log(NT)}}{4\sqrt{2}\left[\sqrt{2C_\zeta p\log((p+1)NT)(C_X\vee C_{Sp}^2C_{Cone}^2)}+C_{Sp}\sqrt{N\vee T}\left(\frac{\sqrt{6C_\zeta}2(1+C_\lambda)}{C_\lambda}+\sqrt{2}\zeta_L\right)\sqrt{\bar{r}}\right]\gamma}.
$$

Then $\mathbb{E}\left[\exp\left(4\tau'\sup_{\Delta_L\in\bar{\mathcal{R}}}\left|\mathbb{G}^0\left(\Delta_{L,it}\right)\right|\right)\right] \le 2\exp\left(\sqrt{\log(NT)}\right) \le 2\sqrt{\exp(\log(NT))} = 2\sqrt{NT}$ for large enough N and T. Substituting it into (S.B.67), we have $\mathbb{E}\left[e^{\tau'\mathcal{C}^1(\gamma)}|W_X,\Omega_1\right] \le 2(p+1)\sqrt{NT}/\gamma$. Let

$$
K_3 = 8\sqrt{2}\left[\sqrt{2C_\zeta p\log((p+1)NT)(C_X\vee C_{Sp}^2C_{Cone}^2)}\right.
$$

$$+ C_{Sp}\sqrt{N \vee T}\left(\frac{\sqrt{6C_\zeta}2(1+C_\lambda)}{C_\lambda} + \sqrt{2}\zeta_L\right)\sqrt{\bar{r}}\Bigg]\gamma\sqrt{\log{(NT)}}.$$

In view of (S.B.66), we thus obtain

$$\mathbb{P}\left(\mathcal{C}^0(\gamma) \geq 2K_3 \middle| \Omega_1\right) = \mathbb{E}\left[\mathbb{P}\left(\mathcal{C}^0(\gamma) \geq 2K_3 \middle| W_X, \Omega_1\right) \middle| \Omega_1\right] \leq \frac{2(p+1)}{\gamma(NT)^{3/2}} \to 0, \qquad \text{(S.B.68)}$$

where the equality is by the law of iterated expectation. Convergence follows the definition of γ and $p = o((N \wedge T)/(\log(NT)\alpha_{NT}^2))$, where the latter is implied by Assumption 3.

Now, recall $s = C_{sup}\sqrt{\log(NT)}\left(\sqrt{p\log((p+1)NT)} \vee \sqrt{\bar{r}(N \vee T)}\right)\gamma$ and let

$$\begin{aligned} C_{sup} =& 192\sqrt{2}\left(\sqrt{2(C_X \vee C_{Sp}^2 C_{Cone}^2)C_\zeta} + C_{Sp}\left(\frac{\sqrt{6C_\zeta}2(1+C_\lambda)}{C_\lambda} + \sqrt{2}\zeta_L\right)\right) \\ =& 192\sqrt{2}\left(2\sqrt{\left(1 \vee \left(\frac{50C_{Sp}^2}{C_\lambda^2 C_{op}^2}\right)\right)C_\zeta C_X} + C_{Sp}\left(\frac{\sqrt{6C_\zeta}2(1+C_\lambda)}{C_\lambda} + \sqrt{2}\zeta_L\right)\right), \end{aligned}$$

where the second equality is by $C_{Cone} = 5\sqrt{2C_X}/C_\lambda C_{op}$. We can see that C_{sup} decreases as C_λ increases, but is bounded away from zero. Note that $s \geq 12\max\{K_1, K_2, 2K_3\}$. Combining equations (S.B.61), (S.B.62), (S.B.63) and (S.B.68), we have

$$\begin{aligned} &\mathbb{P}\left(\mathcal{A}^0(\gamma) \geq \frac{s}{4}\middle|\Omega_1\right) \\ \leq& 3\max\left\{\mathbb{P}\left(\mathcal{B}_1^0(\gamma) \geq \frac{s}{12}\middle|\Omega_1\right), \mathbb{P}\left(\mathcal{B}_2^0(\gamma) \geq \frac{s}{12}\middle|\Omega_1\right), \mathbb{P}\left(\mathcal{C}^0(\gamma) \geq \frac{s}{12}\middle|\Omega_1\right)\right\} \\ \leq& 3\max\left\{\mathbb{P}\left(\mathcal{B}_1^0(\gamma) \geq K_1\middle|\Omega_1\right), \mathbb{P}\left(\mathcal{B}_2^0(\gamma) \geq K_2\middle|\Omega_1\right), \mathbb{P}\left(\mathcal{C}^0(\gamma) \geq 2K_3\middle|\Omega_1\right)\right\} \to 0. \qquad \text{(S.B.69)} \end{aligned}$$

By equations (S.B.59), (S.B.60) and (S.B.69), we obtain $\mathbb{P}(\mathcal{A}(\gamma) \geq s) \to 0$. ■

Proof of Lemma S.B.6

We have

$$\int_0^{w_2}\Big(\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0)\Big)dz - \int_0^{\kappa w_2}\Big(\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0)\Big)dz = \int_{\kappa w_2}^{w_2}\Big(\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0)\Big)dz.$$

For any fixed $w_1 \in \mathbb{R}$, $\mathbb{1}(w_1 \leq z)$ is weakly increasing in z. So, if $w_2 \geq 0$, we have $\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0) \geq 0$ for all $z \in [\kappa w_2, w_2]$ as $\kappa \in (0,1]$; the integral on the right side is nonnegative. If $w_2 < 0$, $\int_{\kappa w_2}^{w_2}\Big(\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0)\Big)dz = \int_{w_2}^{\kappa w_2}\Big(\mathbb{1}(w_1 \leq 0) - \mathbb{1}(w_1 \leq z)\Big)dz$ which is again nonnegative as $\mathbb{1}(w_1 \leq 0) - \mathbb{1}(w_1 \leq z) \geq 0$ for all $z \in [w_2, \kappa w_2]$. ■

Proof of Theorem S.A.1 in Appendix S.A.1

Since $\mathcal{D}^{(2)}$ is a cone, the main argument in the proof of Theorem 2 still holds. Meanwhile, Lemmas 1 and S.B.4 and the upper bound on the penalty difference in equation (S.B.31) do not depend on $\mathcal{D}$ nor on the lower bound obtained in Lemma S.B.2 and Theorem 1. Meanwhile, identification is achieved similar to Theorem 1 once we obtain a lower bound similar to equation (S.B.9) implied by Lemma S.B.2. Therefore, we only need to show that an inequality similar to equation (S.B.9) holds over $(\Delta_\beta, \Delta_L) \in \mathcal{D}^{(2)}$.

With the covariates, recall $\mathcal{D}^{(2)} = \mathbb{R}^p \times \left\{\Delta_L \in \mathbb{R}^{N\times T} : \|\mathcal{P}_\Omega \Delta_L\|_F^2 \geq C_{sm}\|\Delta_L\|_F^2\right\}$. Let $\gamma = C_{error,2}\alpha_{NT}^2\left((1+C_\lambda) \vee \sqrt{\log(NT)}\right)\left(\sqrt{p\log((p+1)NT)/NT} \vee \sqrt{\bar{r}/(N\wedge T)}\right)$ for some new constant $C_{error,2}$. For all $(\Delta_\beta, \Delta_L) \in \mathcal{D}^{(2)}$, all $\|\Delta_\beta\|_F^2 \leq \gamma^2$, and all $u \in \mathcal{U}$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j \Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\right] \\
=&\sum_{i,t}\mathbb{E}\left[\int_0^{\Delta_{L,it}+X_{it}'\Delta_\beta}\Big(F_{V_{it}(u)|W_X}(s) - F_{V_{it}(u)|W_X}(0)\Big)\,ds\right] \\
\geq&\sum_{i,t}\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)\int_0^{\Delta_{L,it}+X_{it}'\Delta_\beta}\Big(F_{V_{it}(u)|W_X}(s) - F_{V_{it}(u)|W_X}(0)\Big)\,ds\right] \\
\geq&\sum_{\{i,t:|\Delta_{L,it}|\leq 2\alpha_{NT}\}}\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)\int_0^{\Delta_{L,it}+X_{it}'\Delta_\beta}\Big(F_{V_{it}(u)|W_X}(s) - F_{V_{it}(u)|W_X}(0)\Big)\,ds\right] \\
&+\sum_{\{i,t:|\Delta_{L,it}|> 2\alpha_{NT}\}}\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)\int_0^{\operatorname{sign}(\Delta_{L,it}+X_{it'}\Delta)|X_{it}'\Delta_\beta|}\Big(F_{V_{it}(u)|W_X}(s) - F_{V_{it}(u)|W_X}(0)\Big)\,ds\right] \\
\geq&\frac{(1\wedge\delta)^2\underline{f}}{18\alpha_{NT}^2}\sum_{i,t}\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)\left(X_{it}'\Delta_\beta + (\mathcal{P}_\Omega\Delta_L)_{it}\right)^2\right] \\
\geq&\frac{(1\wedge\delta)^2\underline{f}}{18\alpha_{NT}^2}\left(0.5\sum_{i,t}\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)\left(X_{it}'\Delta_\beta + \Delta_{L,it}\right)^2\right] - \|\mathcal{P}_{\Omega^\perp}\Delta_L\|_F^2\right) \\
\geq&\frac{(1\wedge\delta)^2\underline{f}}{18\alpha_{NT}^2}\left(0.5\sum_{i,t}\mathbb{E}\left[\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)\left(X_{it}'\Delta_\beta + \Delta_{L,it}\right)^2\right] - (1-C_{sm})\,\|\Delta_L\|_F^2\right) \\
\geq&\frac{0.5C_{min}}{\alpha_{NT}^2}\left(\mathbb{E}\left[\left\|\sum_{j=1}^p X_j\Delta_{\beta,j} + \Delta_L\right\|_F^2\right] - \Big(\varepsilon_{NT}(NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2) + 2(1-C_{sm})\|\Delta_L\|_F^2\Big)\right)
\end{aligned}
$$

where the equality follows the proof of Lemma S.B.2. The first inequality holds because every integral in the summation is nonnegative by Lemma S.B.6. For the second inequality, note that when $\mathbb{1}\left(\sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma\right)$ is true and when $\Delta_{L,it} > 2\alpha_{NT}$, we have

$$
|\Delta_{L,it} + X_{it}'\Delta_\beta| \geq |\Delta_{L,it}| - |X_{it}'\Delta_\beta| \geq |\Delta_{L,it}| - \sqrt{X_{it}'X_{it}}\|\Delta\|_F \geq \alpha_{NT} \geq |X_{it}'\Delta_\beta|,
$$

and thus the inequality holds by Lemma S.B.6. By the definition of $\mathcal{P}_\Omega$, the third inequality holds because now all the upper limits in the two integrals are bounded in magnitude by $3\alpha_{NT}$ given $\|\Delta_\beta\|_F \leq \gamma$. Constants $\underline{f}$ and δ are as in Assumption 2. The fourth and the fifth inequalities are by $\Delta_{L,it} = (\mathcal{P}_\Omega \Delta_L)_{it} + (\mathcal{P}_{\Omega^\perp}\Delta_L)_{it}$ and by the definition of $\mathcal{D}^{(2)}$. The last inequality follows the derivation of equation (S.B.53) in the proof of Lemma S.B.2. We have thus established a result similar to equation (S.B.9), and all the remaining analysis in the paper follows as $C_{sm} \to 1$. ■

S.B.5 Proof of Convergence of Algorithm 1

We start with showing that the inner-loop (equations (A.3) and (A.4)) converges to its global minimum for any fixed $(L^{(k+1)}, H^{(k)})$. Specifically, defining $(\beta^{(k+1)}, V^{(k+1)})$ as

$$(\beta^{(k+1)}, V^{(k+1)}) = \arg\min_{\beta,V} \mathscr{L}(L^{(k+1)}, \beta, V, H^{(k)})$$

where $\mathscr{L}$ is as in equation (A.1) and $(L^{(k+1)}, H^{(k)})$ is given, we show that the accumulation point of the inner loop iterations achieves $\mathscr{L}(L^{(k+1)}, \beta^{(k+1)}, V^{(k+1)}, H^{(k)})$.

Lemma S.B.7. *For any fixed $L^{(k+1)}$ and $H^{(k)}$,*

$$\lim_{l\to\infty} \mathscr{L}(L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)}) = \mathscr{L}(L^{(k+1)}, \beta^{(k+1)}, V^{(k+1)}, H^{(k)}).$$

Proof. For any $(\beta^{(l)}, V^{(l)})$ for all $l \geq 0$, since $\mathscr{L}(L^{(k+1)}, \beta^{(l)}, \cdot, H^{(k)})$ is convex, by definition of the subgradient, we have

$$\begin{aligned}&\mathscr{L}\left(L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)}\right)\\ \leq&\mathscr{L}\left(L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)}\right) - \left\langle V^{(l)} - V^{(l+1)}, \nabla_V \mathscr{L}(L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)})\right\rangle\end{aligned}$$

where $\nabla_V \mathscr{L}(L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)})$ is any subgradient of $\mathscr{L}(L^{(k+1)}, \beta^{(l)}, \cdot, H^{(k)})$ evaluated at $V^{(l+1)}$. By equation (A.7), $0 \in \nabla_V \mathscr{L}(L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)})$, therefore,

$$\mathscr{L}\left(L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)}\right) \leq \mathscr{L}\left(L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)}\right), \forall l.$$

Similarly, since $\mathscr{L}(L^{(k+1)}, \cdot, V^{(l+1)}, H^{(k)})$ is convex and differentiable, and $\beta^{(l+1)}$ is obtained by the first order condition (equation (A.8)),

$$\mathscr{L}\left(L^{(k+1)}, \beta^{(l+1)}, V^{(l+1)}, H^{(k)}\right) \leq \mathscr{L}\left(L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)}\right), \forall l.$$

Combining the above two inequalities, we have the following for any initial value $(\beta^{(0)}, V^{(0)})$,

$$\mathscr{L}(L^{(k+1)}, \beta^{(0)}, V^{(0)}, H^{(k)}) \geq \mathscr{L}(L^{(k+1)}, \beta^{(0)}, V^{(1)}, H^{(k)}) \geq \mathscr{L}(L^{(k+1)}, \beta^{(1)}, V^{(1)}, H^{(k)}) \cdots$$

Therefore, sequence $(\mathscr{L}(L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)}))_l$ is decreasing. By the monotone convergence theorem, $\lim_{l\to\infty} \mathscr{L}(L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)}) \to \mathscr{L}(L^{(k+1)}, \beta^{(k+1)}, V^{(k+1)}, H^{(k)})$ where the limit is the global minimum of $\mathscr{L}(L^{(k+1)}, \cdot, \cdot, H^{(k)})$. ■

Lemma S.B.7 shows that $(\beta^{(l)}, V^{(l)})$ converges to $(\beta^{(k+1)}, V^{(k+1)})$, a global minimizer of $\mathscr{L}(L^{(k+1)}, \cdot, \cdot, H^{(k)})$. From now on, we can treat $(\beta^{(k+1)}, V^{(k+1)})$ as obtained for each k. Now we prove the outer loop also converges. We adapt the proof by Lin et al. (2010) where they prove convergence of an ALM algorithm without covariates and with an increasing μ. Here our μ is fixed to simplify computation. Having a fixed μ requires a new argument especially in the proof of Theorem S.B.1. Define

$$\tilde{H}^{(k+1)} \equiv H^{(k)} - \mu \left(V^{(k)} + \sum_{j=1}^{p} X_j \beta_j^{(k)} + L^{(k+1)} - Y \right). \tag{S.B.70}$$

Combining (A.5) and (S.B.70) and defining $\hat{V} \coloneqq Y - \hat{L} - \sum_{j=1}^{p} X_j \hat{\beta}_j$ where $(\hat{\beta}, \hat{L})$ are our estimator (u is suppressed for the ease of notation), we have

$$\begin{aligned} H^{(k+1)} - H^{(k)} =& \mu \left(Y - V^{(k+1)} - \sum_{j=1}^{p} X_j \beta_j^{(k+1)} - L^{(k+1)} \right) \\ =& \mu \left(\left(\hat{V} - V^{(k+1)} \right) + \sum_{j=1}^{p} X_j \left(\hat{\beta}_j - \beta_j^{(k+1)} \right) + \left(\hat{L} - L^{(k+1)} \right) \right) \end{aligned} \tag{S.B.71}$$

and

$$\tilde{H}^{(k+1)} - H^{(k+1)} = \mu \left(\left(V^{(k+1)} - V^{(k)} \right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)} \right) \right). \tag{S.B.72}$$

Meanwhile, consider the Lagrangian without the penalty:

$$\tilde{\mathscr{L}}(L, \beta, V, H) = \frac{1}{\lambda NT} \boldsymbol{\rho}_u(V) + \|L\|_* + \left\langle H, Y - \sum_{j=1}^{p} X_j \beta_j - L - V \right\rangle. \tag{S.B.73}$$

Note that this is the Lagrangian of the minimization problem that defines our estimator $(\hat{\beta}, \hat{L})$. Let $\hat{H}$ be the Lagrangian multiplier when $(L, \beta, V) = (\hat{L}, \hat{\beta}, \hat{V})$. Then we have the following lemma.

Lemma S.B.8. *For all $j = 1, \dots, p$ and for all $k \geq 1$, we have*

$$\left\langle \hat{H}, X_j \right\rangle = \left\langle H^{(k+1)}, X_j \right\rangle = 0. \tag{S.B.74}$$

Proof. The first order condition of equation (S.B.73) with respect to β_j is $\left\langle \hat{H}, X_j \right\rangle = 0$ for all $j = 1, \dots, p$. Similarly, equation (A.4) at $(\beta^{(k+1)}, V^{(k+1)})$ is

$$0 = \left\langle H^{(k)}, X_j \right\rangle + \mu \left\langle Y - \sum_{j=1}^{p} X_j \beta_j^{(k+1)} - L^{(k+1)} - V^{(k+1)}, X_j \right\rangle, \forall j = 1, ..., p. \tag{S.B.75}$$

Together with (A.5), we have $\left\langle H^{(k+1)}, X_j \right\rangle = 0$ for all $j = 1, ..., p$. ■

Lemma S.B.9. *For $k \geq 2$, the following holds:*

$$\begin{aligned}
&\left\| \left(V^{(k+1)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \hat{\beta}_j\right) \right\|_F^2 + \mu^{-2} \|H^{(k+1)} - \hat{H}\|_F^2 \\
=& \left(\left\| \left(V^{(k)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k)} - \hat{\beta}_j\right) \right\|_F^2 - \left\| \left(V^{(k+1)} - V^{(k)}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)}\right) \right\|_F^2 \right) \\
&+ \mu^{-2} \left(\left\| H^{(k)} - \hat{H} \right\|_F^2 - \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 \right) \\
&- 2\mu^{-1} \left(\left\langle V^{(k+1)} - V^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle + \left\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \right\rangle + \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle \right).
\end{aligned}$$

Proof. By Lemma S.B.8, and by equations (S.B.71) and (S.B.72), we have

$$\begin{aligned}
&\mu^{-1} \left\langle H^{(k+1)} - H^{(k)}, H^{(k+1)} - \hat{H} \right\rangle \\
=& - \left\langle V^{(k+1)} - \hat{V} + \sum_{j=1}^{p} X_j (\beta_j^{(k+1)} - \hat{\beta}_j), H^{(k+1)} - \hat{H} \right\rangle - \left\langle L^{(k+1)} - \hat{L}, H^{(k+1)} - \hat{H} \right\rangle \\
=& - \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle - \left\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \right\rangle \\
&+ \mu \left\langle L^{(k+1)} - \hat{L}, \left(V^{(k+1)} - V^{(k)}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)}\right) \right\rangle.
\end{aligned} \tag{S.B.76}$$

Therefore,

$$\begin{aligned}
&\left\| \left(V^{(k+1)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \hat{\beta}_j\right) \right\|_F^2 + \mu^{-2} \|H^{(k+1)} - \hat{H}\|_F^2 \\
=& \left(\left\| \left(V^{(k)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k)} - \hat{\beta}_j\right) \right\|_F^2 - \left\| \left(V^{(k+1)} - V^{(k)}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)}\right) \right\|_F^2 \right)
\end{aligned}$$

$$
\begin{aligned}
&+ 2\left\langle \left(V^{(k+1)} - V^{(k)}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)}\right), \left(V^{(k+1)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \hat{\beta}_j\right) \right\rangle \\
&+ \mu^{-2} \left(\|H^{(k)} - \hat{H}\|_F^2 - \|H^{(k+1)} - H^{(k)}\|_F^2 + 2\left\langle H^{(k+1)} - H^{(k)}, H^{(k+1)} - \hat{H} \right\rangle \right) \\
= &\left(\left\| \left(V^{(k)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k)} - \hat{\beta}_j\right) \right\|_F^2 - \left\| \left(V^{(k+1)} - V^{(k)}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)}\right) \right\|_F^2 \right) \\
&+ \mu^{-2} \left(\|H^{(k)} - \hat{H}\|_F^2 - \|H^{(k+1)} - H^{(k)}\|_F^2 \right) \\
&- 2\mu^{-1} \left\langle \left(V^{(k+1)} - V^{(k)}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)}\right), H^{(k+1)} - H^{(k)} \right\rangle \\
&- 2\mu^{-1} \left(\left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle + \left\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \right\rangle \right) \\
= &\left(\left\| \left(V^{(k)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k)} - \hat{\beta}_j\right) \right\|_F^2 - \left\| \left(V^{(k+1)} - V^{(k)}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \beta_j^{(k)}\right) \right\|_F^2 \right) \\
&+ \mu^{-2} \left(\left\| H^{(k)} - \hat{H} \right\|_F^2 - \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 \right) \\
&- 2\mu^{-1} \left(\left\langle V^{(k+1)} - V^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle + \left\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \right\rangle + \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle \right),
\end{aligned}
$$

where the first equality is due to the fact that $\|A - B\|_F^2 = \|C - B\|_F^2 - \|A - C\|_F^2 + 2\langle A - C, A - B\rangle$ for all $N \times T$ matrices A, B and C. The second equality is by equations (S.B.71) and (S.B.76). The last equality is by Lemma S.B.8. ■

Next we are to show that the three inner products on the right hand side of the last equality above are all nonnegative. That proves the left hand side of the first equality above is a decreasing sequence in k. We need the following lemmas.

Lemma S.B.10 (Lemma 3 in Lin et al. (2010))**.** *If f is a convex function, then $\langle x_1 - x_2, g_1 - g_2\rangle \geq 0$, $\forall g_i \in \nabla f(x_i), i = 1, 2$, where ∇f is a subgradient of f.*

Lemma S.B.11.

$$
\begin{aligned}
\lim_{k\to\infty} \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 &= 0, \\
\lim_{k\to\infty} \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle &= 0, \\
\lim_{k\to\infty} \left\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \right\rangle &= 0.
\end{aligned}
$$

Proof. By the optimality of $(\hat{L}, \hat{\beta}, \hat{V}, \hat{H})$ to the unpenalized Lagrangian (S.B.73), we have

$$
\hat{H} \in \nabla\|\hat{L}\|_*, \qquad \hat{H} \in \frac{1}{\lambda NT} \nabla \boldsymbol{\rho}_u(\hat{V}). \tag{S.B.77}
$$

By equations (A.2), (A.3), (A.5) and (S.B.70) where in equation (A.3), $\beta^{(l)}$ and $V^{(l+1)}$ are replaced with $\beta^{(k+1)}$ and $V^{(k+1)}$, we have

$$\tilde{H}^{(k+1)} \in \nabla \|L^{(k+1)}\|_*, \qquad H^{(k+1)} \in \frac{1}{\lambda NT} \nabla \boldsymbol{\rho}_u(V^{(k+1)}). \tag{S.B.78}$$

Therefore, by Lemma S.B.10 we have the following inequalities:

$$\begin{aligned} \left\langle V^{(k+1)} - V^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle &\geq 0, \\ \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle &\geq 0, \\ \left\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \right\rangle &\geq 0. \end{aligned}$$

Now by the nonnegativeness of the three inner products and by Lemma S.B.9, we have that $\Psi^{(k)} := \left\| \left(V^{(k+1)} - \hat{V}\right) + \sum_{j=1}^{p} X_j \left(\beta_j^{(k+1)} - \hat{\beta}_j\right) \right\|_F^2 + \mu^{-2} \|H^{(k+1)} - \hat{H}\|_F^2$ is decreasing in k. Since $\Psi^{(k)}$ is lower bounded by zero, it is convergent by the monotone convergence theorem. Therefore, by Lemma S.B.9 and by the nonnegativeness of the three inner products, we have

$$\begin{aligned} 0 \leq & \mu^{-2} \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 + \mu^{-1} \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle + \mu^{-1} \left\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \right\rangle \\ \leq & \Psi^{(k)} - \Psi^{(k+1)} \to 0. \end{aligned}$$

Hence, the desired results obtain since $\left\| H^{(k+1)} - H^{(k)} \right\|_F^2$, $\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \rangle$ and $\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \rangle$ are all nonnegative. ■

Theorem S.B.1. *Algorithm 1 converges to the global minimum of the objective function (2.3):*

$$\lim_{k \to \infty} \left\| Y - L^{(k)} - \sum_{j=1}^{p} X_j \beta_j^{(k)} - V^{(k)} \right\|_F^2 = 0, \tag{S.B.79}$$

$$\lim_{k \to \infty} \frac{1}{\lambda NT} \boldsymbol{\rho}_u \left(Y - \sum_{j=1}^{p} X_j \beta_j^{(k)} - L^{(k)} \right) + \left\| L^{(k)} \right\|_* = \frac{1}{\lambda NT} \boldsymbol{\rho}_u \left(Y - \hat{L} - \sum_{j=1}^{p} X_j \hat{\beta}_j \right) + \|\hat{L}\|_*. \tag{S.B.80}$$

Proof. Since $\|H^{(k)} - H^{(k-1)}\|_F^2 \to 0$ as $k \to \infty$ by Lemma S.B.11, under equation (S.B.71),

$$\left\| Y - L^{(k)} - \sum_{j=1}^{p} X_j \beta_j^{(k)} - V^{(k)} \right\|_F^2 = \mu^{-1} \left\| H^{(k)} - H^{(k-1)} \right\|_F^2 \to 0. \tag{S.B.81}$$

Therefore, constraint $Y = L^{(k)} + \sum_{j=1}^{p} X_j \beta_j^{(k)} + V^{(k)}$ holds in the limit and equation (S.B.79) is proved.

Recall that $\tilde{H}^{(k)} \in \nabla\|L^{(k)}\|_*$ and that $H^{(k)} \in \frac{1}{\lambda NT}\nabla\boldsymbol{\rho}_u(V^{(k)})$ by equation (S.B.78). By the definition of subgradient,

$$\begin{aligned}
&\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(V^{(k)}\right)+\left\|L^{(k)}\right\|_*\\
\leq&\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*-\left\langle\hat{L}-L^{(k)},\tilde{H}^{(k)}\right\rangle-\left\langle\hat{V}-V^{(k)},H^{(k)}\right\rangle\\
=&\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*+\left\langle\hat{L}-L^{(k)},\hat{H}-\tilde{H}^{(k)}\right\rangle+\left\langle\hat{V}-V^{(k)},\hat{H}-H^{(k)}\right\rangle\\
&-\left\langle\hat{L}+\hat{V}-L^{(k)}-V^{(k)},\hat{H}\right\rangle\\
=&\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*+\left\langle\hat{L}-L^{(k)},\hat{H}-\tilde{H}^{(k)}\right\rangle+\left\langle\hat{V}-V^{(k)},\hat{H}-H^{(k)}\right\rangle\\
&-\left\langle Y-\sum_{j=1}^{p}X_j\hat{\beta}_j-L^{(k)}-V^{(k)},\hat{H}\right\rangle\\
=&\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*+\left\langle\hat{L}-L^{(k)},\hat{H}-\tilde{H}^{(k)}\right\rangle+\left\langle\hat{V}-V^{(k)},\hat{H}-H^{(k)}\right\rangle\\
&-\left\langle Y-\sum_{j=1}^{p}X_j\beta_j^{(k)}-L^{(k)}-V^{(k)},\hat{H}\right\rangle\\
\leq&\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*+\left\langle\hat{L}-L^{(k)},\hat{H}-\tilde{H}^{(k)}\right\rangle+\left\langle\hat{V}-V^{(k)},\hat{H}-H^{(k)}\right\rangle\\
&+\left\|\hat{H}\right\|_F\cdot\left\|Y-\sum_{j=1}^{p}X_j\beta_j^{(k)}-L^{(k)}-V^{(k)}\right\|_F\\
=:&\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*+a_k.
\end{aligned}\tag{S.B.82}$$

The first equality is elementary. The second equality is by $Y=\sum_{j=1}^{p}X_j\hat{\beta}_j+\hat{L}+\hat{V}$. The third equality is by $\langle X_j,\hat{H}\rangle=0$ by Lemma S.B.8. The last inequality is by the Cauchy-Schwarz inequality. By Lemma S.B.11 and by equation (S.B.81), $a_k\to 0$ as $k\to\infty$ since $\|\hat{H}\|_F$ stays constant across k.

On the other hand, let $\tilde{V}^{(k)}=Y-L^{(k)}-\sum_{j=1}^{p}X_j\beta_j^{(k)}$. By construction, $(L^{(k)},\beta^{(k)},\tilde{V}^{(k)})$ satisfy the constraint. So,

$$\left\|\tilde{V}^{(k)}-V^{(k)}\right\|_F=\left\|Y-L^{(k)}-\sum_{j=1}^{p}X_j\beta_j^{(k)}-V^{(k)}\right\|_F\to 0$$

where convergence is by equation (S.B.81). So $\tilde{V}^{(k)}-V^{(k)}\to 0$ componentwise as $k\to\infty$. Therefore, $b_k:=\|\tilde{V}^{(k)}-V^{(k)}\|_1/(\lambda NT)$ converges to 0 as $k\to\infty$. By contractivity of the check function, we have

$$\left|\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(V^{(k)}\right)+\left\|L^{(k)}\right\|_*-\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\tilde{V}^{(k)}\right)-\left\|L^{(k)}\right\|_*\right|\leq b_k\to 0 \tag{S.B.83}$$

Therefore,

$$\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(V^{(k)}\right)+\left\|L^{(k)}\right\|_*\geq\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\tilde{V}^{(k)}\right)+\left\|L^{(k)}\right\|_*-b_k\geq\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*-b_k \tag{S.B.84}$$

where the first inequality is by equation (S.B.83) and the second inequality is by the definition of the estimator which says for all (L,V) such that there exists a β satisfying the constraint $L+V+\sum_{j=1}^{p}X_j\beta_j=Y$, $\boldsymbol{\rho}_u(V)/(\lambda NT)+\|L\|_*$ is minimized at $(\hat{V},\hat{L})$. Combining equations (S.B.82) and (S.B.84), we have

$$\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*-b_k\leq\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(V^{(k)}\right)+\left\|L^{(k)}\right\|_*\leq\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(\hat{V}\right)+\left\|\hat{L}\right\|_*+a_k.$$

Therefore, by $\hat{V}=Y-\sum_{j=1}^{p}X_j\hat{\beta}-\hat{L}$, the following holds by $a_k,b_k\to 0$:

$$\lim_{k\to\infty}\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(V^{(k)}\right)+\left\|L^{(k)}\right\|_*=\frac{1}{\lambda NT}\boldsymbol{\rho}_u\left(Y-\sum_{j=1}^{p}X_j\hat{\beta}-\hat{L}\right)+\left\|\hat{L}\right\|_*. \tag{S.B.85}$$

In view of equation (S.B.83) and $\tilde{V}^{(k)}=Y-L^{(k)}-\sum_{j=1}^{p}X_j\beta_j^{(k)}$, equation (S.B.85) thus implies equation (S.B.80). ∎

References

Ando, T. and J. Bai (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association 115*(529), 266–279.

Belloni, A., M. Chen, O. H. M. Padilla, and Z. K. Wang (2023). High-dimensional latent panel quantile regression with an application to asset pricing. *The Annals of Statistics 51*(1), 96 – 121.

Belloni, A. and V. Chernozhukov (2011). ℓ1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*(1), 82–130.

Candès, E. J. and B. Recht (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics 9*(6), 717–772.

Chao, S.-K., W. K. Härdle, and M. Yuan (2021). Factorisable multitask quantile regression. *Econometric Theory 37*(4), 794–816.

Chen, L., J. J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics 26*(2), 755–770.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: Isoperimetry and Processes.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer-Verlag.

Lin, Z., M. Chen, and Y. Ma (2010). The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

Moon, H. R. and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica 83*(4), 1543–1579.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction With Applications in Data Science*, Volume 47. Cambridge University Press.