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 $\|\tilde{L}(u)\|_\infty \leq \alpha_{NT}$ 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 $\|\Delta_L\|^2_F$ multiplied by some constant for all $\Delta_L \in \mathcal{D} := \{\Delta_L \in \mathbb{R}^{N \times T} : \|\Delta_L\|_{\infty} \leq 2\alpha_{NT}\}$ (see equation (4.6)):

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

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 $\|L_0(u)\|_{\infty} \leq \alpha_{NT}$.

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 $1(|\Delta_{L,it}| \leq 2\alpha_{NT}) \cdot \Delta_{L,it}$. Let $\mathcal{P}_\Omega \perp \Delta_L := \Delta - \mathcal{P}_\Omega \Delta_L$. By construction, $\|\Delta_L\|^2_F = \|\mathcal{P}_\Omega \Delta_L\|^2_F + \|\mathcal{P}_\Omega \perp \Delta_L\|^2_F$. Let $\|\mathcal{P}_\Omega \Delta_L\|^2_F = C_{sm}\|\Delta_L\|^2_F$ 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

$$\sum_{i,t} \int_0^{\Delta_{L,it}} (F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0)) \, ds = \sum_{\{i,t:|\Delta_{L,it}| \leq 2\alpha_{NT}\}} \int_0^{\Delta_{L,it}} (F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0)) \, ds$$

$$+ \sum_{\{i,t:|\Delta_{L,it}| > 2\alpha_{NT}\}} \int_0^{\Delta_{L,it}} (F_{V_{it}(u)}(s) - F_{V_{it}(u)}(0)) \, ds.$$ 

For the first sum on the right side, we can lower bound it by $C_{min}\|\mathcal{P}_\Omega \Delta_L\|^2_F/\alpha_{NT}^2$ 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 $\|\mathcal{P}_\Omega \perp \Delta_L\|^2_F$ is of a nonnegligible proportion of $\|\Delta_L\|^2_F$, we can still lower bound $(S.A.1)$ be $\|\Delta\|^2_F$ 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\(^1\)

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

\(^1\)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 $\|\Delta_L\|_F \leq \sqrt{NT\gamma}$ 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 $\|\Delta_L\|_F = \sqrt{NT\gamma}$. 

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

$$C_{\text{min}}\|P_{\Omega}\Delta L\|_F^2/\alpha_{NT}^2 + 0 \geq C_{\text{sm}}C_{\text{min}}\|\Delta L\|_F^2/\alpha_{NT}^2.$$  

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

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

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

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

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

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

**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 \( u \in U \), 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 \( D \) 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 \( \hat{\Delta}_L(u) \in D \) for all \( u \in U \).

- **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 \( \hat{\Delta}_L(u) \in D^{(2)} \) for all \( u \in U \) 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 \( f_S \) such that the density \( f_{V_{it}(u)}(s) \geq f_S > 0 \) 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 \( \| L_0(u) \|_{\infty} \leq \alpha_{NT} \) and \( \| L \|_{\infty} \leq \alpha_{NT} \) are still imposed. To see this, by \( |\Delta_{L,it}| \leq 2\alpha_{NT} \), their assumption implies that there exists a constant \( f_{\alpha_{NT}} > 0 \) such that (S.A.1) is lower bounded by \( f_{\alpha_{NT}} \sum_{i,t} \Delta_{L,it}^2/2 \) by directly applying first-order Taylor expansion. Similar to our Approach 1, the lower bound also depends on \( \alpha_{NT} \) via \( f_{\alpha_{NT}} \).

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 \( f \) 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. \( \alpha_{NT} \) is fixed, not \( (N,T) \)-dependent.
\[ \partial f_{V_{i}(u)}(v)/\partial v \text{ is continuous and bounded in absolute value by } \tilde{f}' \text{ uniformly in } v, i, \text{ and } t. \]

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\(^3\)

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

and the constants \( f \) and \( \tilde{f}' \) 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:

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

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 \( \nu_{NT} \) 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 \( o(\nu_{NT}^2/\delta_{NT}^3) \) 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 \( f_{V_{i}(u)} \) and ii) to restrict the \textit{magnitude} of large elements in \( \hat{\Delta}_L(u) \) (Approach 1) or both the \textit{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 \( \| L_0(u) \|_\infty \leq \alpha_{NT} \),

\[ \text{More specifically, in a similar way to footnote 1, the assumption is imposed on all } \Delta_L \in \mathcal{R}_u; \text{ it basically requires } \gamma/4 \leq 3\| \Delta_L \|^3_F/(8\sqrt{NT} \tilde{f}') \sum_{i,t} |\Delta_{L,it}|^3 \text{ (supposing there are no covariates). Under this assumption, } \Delta_L \in \mathcal{D}^{(3)} \text{ as long as } \Delta_L \in \mathcal{R}_u \text{ with } \| \Delta_L \|_F \leq \sqrt{NT} \gamma. \]
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 $W$ for simplicity and $V(u) := Y - q_Y|W_X(u)$. Let $\nabla \rho_u(V(u))$ be an $N \times T$ subgradient matrix of $\rho_u(\cdot)$ evaluated at $V(u)$. Note that for check function $\rho_u(\cdot)$, $\nabla \rho_u(0)$ is not unique due to nonsmoothness of $\rho_u$ 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 $\nabla \rho_u(V(u))$ is

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

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 $V(u)$. We introduce the following lemma for $\nabla \rho_u(V(u))$. 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:

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

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

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:

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

---

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 \( \frac{1}{NT} \left[ \rho_u \left( V(u) - \sum_{j=1}^{\beta,j} X_j \hat{\Delta}_{\beta,j}(u) - \hat{\Delta}_L(u) \right) - \rho_u(V(u)) \right] / NT \). With probability one,

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

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, \( \left| \langle A, B \rangle \right| \leq \| A \| \cdot \| B \|_* \). The penultimate inequality is by inequalities (S.B.1) and (S.B.2) in Lemma S.B.1.

Next, consider \( \lambda \left( \| \hat{L}(u) \|_* - \| L_0(u) \|_* \right) \). 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,

\[
\| \hat{L}(u) \|_* - \| L_0(u) \|_* = \| \mathcal{P}_{\Phi(u)} L_0(u) + \mathcal{P}_{\Phi(u)} \hat{\Delta}_L(u) \|_* + \| \mathcal{P}_{\Phi(u)^\perp} \hat{\Delta}_L(u) \|_* - \| \mathcal{P}_{\Phi(u)} L_0(u) \|_* \\
\geq \| \mathcal{P}_{\Phi(u)^\perp} \hat{\Delta}_L(u) \|_* - \| \mathcal{P}_{\Phi(u)^\perp} \hat{\Delta}_L(u) \|_* \tag{S.B.5}
\]

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

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

So, we have

\[
\sup_{u \in \mathcal{U}} \left( \| \mathcal{P}_{\Phi(u)^\perp} \hat{\Delta}_L(u) \|_* - \frac{NT \lambda + C_{op} \sqrt{N \vee T}}{NT \lambda - C_{op} \sqrt{N \vee T}} \| \mathcal{P}_{\Phi(u)} \hat{\Delta}_L(u) \|_* \right) \leq 0.
\]
\[-\frac{5}{\sqrt{2}C_\text{opt} \log((p+1)NT)}}{NT \lambda - C_\text{opt} \sqrt{N \vee T}} \| \hat{\Delta}_\beta(u) \|_F \leq 0\]

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 \( \lambda = (1 + C_\lambda) C_\text{opt} \sqrt{N \vee T}/NT \) into it.

\section*{S.B.2 Proof of the Results in Section 4}

\textbf{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

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

where \( \text{Tr} \) 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

\[
\mathbb{E} \left\| \mathcal{P}_{\Phi(u)^\perp} \left( \sum_{j=1}^{p} X_j \tau_j \right) \right\|_F^2 = \mathbb{E} \left\| M_{\Lambda_0}(u) \left( \sum_{j=1}^{p} X_j \tau_j \right) M_{F_0}(u) \right\|_F^2 \\
= \tau' \mathbb{E} (x' (M_{F_0}(u) \otimes M_{\Lambda_0}(u)) x) \tau,
\]

where the first equality is by Remark 6. Similarly,

\[
\mathbb{E} \left\| \mathcal{P}_{\Phi(u)} \left( \sum_{j=1}^{p} X_j \tau_j \right) \right\|_F^2 = \mathbb{E} \left\| \sum_{j=1}^{p} X_j \tau_j \right\|_F^2 - \mathbb{E} \left\| \mathcal{P}_{\Phi(u)^\perp} \left( \sum_{j=1}^{p} X_j \tau_j \right) \right\|_F^2 \\
= \tau' \mathbb{E} (x' x) \tau - \tau' \mathbb{E} (x' (M_{F_0}(u) \otimes M_{\Lambda_0}(u)) x) \tau.
\]

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 \( E \left\| \sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L \right\|_{F}^2 \) (see Lemma S.B.2 for details). Then it is sufficient to show that \( E \left\| \sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L \right\|_{F}^2 \geq C_{\text{RSC}} (NT\|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2) \) for all \((\Delta_\beta, \Delta_L) \in \mathcal{R}_u\) for some \(C_{\text{RSC}} > 0\) and Theorem 1 follows. We distinguish two cases. The first case is \( r(u) = 0 \). In this case, for any \( N \times T \) matrix \( M, P_{\Phi(u)} M = 0 \). Therefore, if \((\Delta_\beta, \Delta_L) \in \mathcal{R}_u\), it is necessary that \( \|\Delta_L\|_F^2 = \|P_{\Phi(u)^\perp}\Delta_L\|_F^2 \leq \kappa_2(\lambda)^2 \|\Delta_\beta\|_F^2 \) 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, \( \|\Delta_L\|_F^2 = o(N^2 \wedge T^2) \|\Delta_\beta\|_F^2 \). On the other hand, Assumption 4 ii) implies that \( E\left\| \sum_{j=1}^{p} X_j \Delta_{\beta,j} \right\|_F^2 \) is no smaller than \( \sigma_{\text{min}}^2 NT\|\Delta_\beta\|_F^2 \). Therefore, \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L\|_F^2 \) is greater than \( \|\Delta_L\|_F^2 \) in order. So, the quantity \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L\|_F^2 \) must be bounded from below by a large fraction of \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\|_F^2 \), and consequently by (still a large fraction of) \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\|_F^2 + \|\Delta_L\|_F^2 \) 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 \( \|P_{\Phi(u)^\perp}\Delta_L\|_F \leq \sqrt{3r(u)\kappa_1(\lambda)}\|P_{\Phi(u)}\Delta_L\|_F + \kappa_2(\lambda)\|\Delta_\beta\|_F \). Similar to the \( r(u) = 0 \) case, a simple case is when \( \sqrt{3r(u)\kappa_1(\lambda)}\|P_{\Phi(u)}\Delta_L\|_F \leq \kappa_2(\lambda)\|\Delta_\beta\|_F \). In this case, the order of \( \|\Delta_L\|_F \) is bounded by \( \kappa_2(\lambda)\|\Delta_\beta\|_F \) because both the order of \( \|P_{\Phi(u)^\perp}\Delta_L\|_F \) and of \( \|P_{\Phi(u)}\Delta_L\|_F \) are bounded by it. Again, this is dominated by the order of \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\|_F^2 \) following the argument as the \( r(u) = 0 \) case.

The nontrivial case is when \( r(u) \geq 1 \) and \( \sqrt{3r(u)\kappa_1(\lambda)}\|P_{\Phi(u)}\Delta_L\|_F > \kappa_2(\lambda)\|\Delta_\beta\|_F \). In this case, it is possible that \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j}\|_F^2 \) and \( \|\Delta_L\|_F^2 \) are of the same order, and consequently, without further restriction, the quantity of interest \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L\|_F^2 \) can reach 0 even when \( \Delta_L \) and \( \Delta_\beta \) 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 \( E\|\sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L\|_F^2 \) is bounded from below by

\[
\left( c_1(u) \sqrt{E\|P_{\Phi(u)^\perp}\left(\sum_{j=1}^{p} X_j \Delta_{\beta,j}\right)\|_F^2} \right)^2 \left( \frac{\kappa_2(\lambda)}{\sqrt{3r(u)\kappa_1(\lambda)}} + \|P_{\Phi(u)}\Delta_L\|_F \right)^2
\]
+ \left( \sqrt{\mathbb{E} \left\| 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{3\rho_c(u)\kappa_1(\lambda)}} + \| P_{\Phi(u)\Delta_L \perp} \|_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 $\varepsilon_{NT} := 2 \left( \max_{i,t} \mathbb{E} \left( 1 \left( \sqrt{X'_i X_{it}} > \alpha_{NT}/\gamma \right) \cdot X'_i X_{it} \right) \vee \max_{i,t} \mathbb{P} \left( \sqrt{X'_i X_{it}} > \alpha_{NT}/\gamma \right) \right)$. Under Assumption 2, if $\alpha_{NT} \geq 1$, then the following inequality holds for any fixed $N$ and $T$:

\[
\inf_{\Delta_L} \left( \mathbb{E} \left( \rho_u(V(u) - \sum_{j=1}^{p} X_j \Delta_{\beta,j} - \Delta_L) - \rho_u(V(u)) \right) \right)
- (1 \wedge \delta)^2 f \left( \mathbb{E} \left\| \sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L \right\|_F^2 - \varepsilon_{NT} \left( NT \| \Delta_{\beta} \|_F^2 + \| \Delta_L \|_F^2 \right) \right) \\
\geq 0. \quad \text{(S.B.8)}
\]

Sequence $\varepsilon_{NT}$ 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:

\[
\inf_{\Delta_L \in \mathcal{D}} \left( \mathbb{E} \left( \rho_u(V(u) - \sum_{j=1}^{p} X_j \Delta_{\beta,j} - \Delta_L) - \rho_u(V(u)) \right) \right)
- \frac{C_{\alpha_{NT}}^2}{\alpha_{NT}^2} \left( \mathbb{E} \left\| \sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L \right\|_F^2 - \varepsilon_{NT} \left( NT \| \Delta_{\beta} \|_F^2 + \| \Delta_L \|_F^2 \right) \right) \geq 0, \quad \text{(S.B.9)}
\]

where $C_{\alpha_{NT}} := (1 \wedge \delta)^2 f/18$, $f$ and $\delta$ are defined in Assumption 2, and $\gamma$ 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 $\kappa_1$ and

10
for all \( \kappa_2 \). 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 \( \kappa_3 \) has the same order as \( \kappa_2 \) 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,

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

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 \( \kappa_2 \) 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 \( u \in \{ u \in \mathcal{U} : r(u) = 0 \} \),

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

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 \( T > T_1 \),

\[
\inf_{u \in \{ u \in \mathcal{U} : r(u) = 0 \}} \left( \mathbb{E} \left\| \sum_{j=1}^{p} X_j \Delta_{\beta,j} + \Delta_L \right\|_F^2 - (1 - \varepsilon_1) \sigma_{\min}^2 \wedge 1 \left( NT \|\Delta_\beta\|_F^2 + \|\Delta_L\|_F^2 \right) \right) \geq 0. \tag{S.B.13}
\]
Case 2.1. \( r(u) \geq 1 \) and \( \|P_{\Phi(u)} \Delta L\|_F \leq \kappa_3 \|\Delta_\beta\|_F \). By the definition of \( R_u \), \( \|P_{\Phi(u)} \Delta L\|_F \leq \sqrt{3r(u)\kappa_1 |P_{\Phi(u)} \Delta L|}_F + \kappa_2 \|\Delta_\beta\|_F \) if \((\Delta_\beta, \Delta L) \in R_u\). By the definition of \( \kappa_3 \), we then have

\[
\|P_{\Phi(u)} \Delta L\|_F \leq \sqrt{3r(u)\kappa_1 \left(\|P_{\Phi(u)} \Delta L\|_F + \kappa_3 \|\Delta_\beta\|_F\right)} \leq 2\sqrt{3r(u)\kappa_1 \kappa_3 \|\Delta_\beta\|_F}.
\]

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 \( o((N \wedge T)^2) \|\Delta_\beta\|_F^2 \) 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 \( T > T_2 \):

\[
\inf_{u \in \{u \in U: r(u) \geq 1\} \atop (\Delta_\beta, \Delta L) \in R_u} \mathbb{E} \left( \left\| \sum_{j=1}^p X_j \Delta_{\beta,j} + \Delta_L \right\|_F^2 - \left(1 - \varepsilon_1 \sigma_{min}^2 \wedge 1\right) (NT \|\Delta_\beta\|_F^2 + \|\Delta L\|_F^2) \right) \geq 0.
\]  

(S.B.14)

Case 2.2. \( r(u) \geq 1 \) and \( \|P_{\Phi(u)} \Delta L\|_F > \kappa_3 \|\Delta_\beta\|_F \). Let \( M(u) := \sqrt{\mathbb{E} \left\|P_{\Phi(u)} \left(\sum_{j=1}^p X_j \Delta_{\beta,j}\right)\right\|_F^2} \) and \( M(u)^\perp := \sqrt{\mathbb{E} \left\|P_{\Phi(u)} \perp \left(\sum_{j=1}^p X_j \Delta_{\beta,j}\right)\right\|_F^2} \). 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 \( u \in U \), then there exist \( N_3 \) and \( T_3 \) such that the following holds for all \( N > N_3 \) and \( T > T_3 \):

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

**Remark S.B.2.** Recall that for any \((\Delta_\beta, \Delta L) \in R_u\), by the definition of the cone, we can verify that the following holds:

\[
\|P_{\Phi(u)} \perp \Delta L\|_F \leq \left(\sqrt{3r(u)\kappa_1(\lambda)} \|P_{\Phi(u)} \Delta L\|_F + \kappa_2(\lambda) \|\Delta_\beta\|_F\right).
\]

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 \((\Delta_\beta, \Delta L)\) 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 \( u \in U \) such that \( r(u) \geq 1 \), Lemma S.B.3 implies that there exists a \( c_1(u) \in [0, 1/\left(\sqrt{3r(u)\kappa_1 C_{\Phi X}}\right)] \) such that \((c_1(u) \text{ may depend on } \Delta_L \text{ and } \Delta_\beta)\)

\[
c_1(u)M(u)^\perp = M(u) + \kappa_3 \|\Delta_\beta\|_F \implies M(u) = c_1(u)M(u)^\perp - \kappa_3 \|\Delta_\beta\|_F \text{,} \quad (S.B.15)
\]
where \( c_1(u) = 0 \) if and only if \( \Delta_\beta = 0 \). Since \( \kappa_1 = (2 + C_\lambda)/C_\lambda > 1, \) \( 1 \leq r(u) \leq \tilde{r}, \) and \( C_{\Phi_X} > 1, \) we have
\[
0 \leq c_1(u) \leq \frac{1}{\sqrt[3]{3r(u)\kappa_1 C_{\Phi_X}}} < \frac{1}{\sqrt[3]{3r(u)\kappa_1}} < 1,
\] (S.B.16)
for all \( u \in \{u \in U : r(u) \geq 1\} \) and all \( (\Delta_\beta, \Delta_L) \).

Similarly, for \( (\Delta_\beta, \Delta_L) \in \mathcal{R}_u, \) since
\[
\left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F \leq \sqrt[3]{3r(u)\kappa_1} \left( \left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F + \kappa_3 \|\Delta_\beta\|_F \right)
\]
by the definition of \( \mathcal{R}_u, \) there exists a \( c_2(u) \in [0, \sqrt[3]{3r(u)\kappa_1}] \) that may depend on \( \Delta_L \) and \( \Delta_\beta \) as well such that
\[
\left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F = c_2(u) \left( \left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F + \kappa_3 \|\Delta_\beta\|_F \right).
\] (S.B.17)

Then we have
\[
0 \leq c_2(u) \leq \sqrt[3]{3r(u)\kappa_1} = \frac{\sqrt[3]{3}(2 + C_\lambda)}{C_\lambda},
\]
(S.B.18)
\[
0 \leq c_1(u)c_2(u) \leq \frac{1}{C_{\Phi_X}} < 1.
\] (S.B.19)
for all \( u \in \{u \in U : r(u) \geq 1\} \) and all \( (\Delta_\beta, \Delta_L) \).

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 \( u \in \{u \in U : r(u) \geq 1\} \):
\[
\mathbb{E} \left\| \sum_{j=1}^p X_j \Delta_{\beta,j} + \Delta_L \right\|_F^2
= \mathbb{E} \left\| \mathcal{P}_{\Phi(u)} \left( \sum_{j=1}^p X_j \Delta_{\beta,j} \right) + \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F^2 + \mathbb{E} \left\| \mathcal{P}_{\Phi(u)} \left( \sum_{j=1}^p X_j \Delta_{\beta,j} \right) + \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F^2
\geq \left( M(u) - \left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F \right)^2 + \left( M(u) - \left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F \right)^2
= \left( c_1(u)M(u) - \left( \kappa_3 \|\Delta_\beta\|_F + \left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F \right) \right)^2 + \left( M(u) - c_2(u) \left( \kappa_3 \|\Delta_\beta\|_F + \left\| \mathcal{P}_{\Phi(u)} \Delta_L \right\|_F \right) \right)^2.
\] (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 \( u \in \{u \in U : r(u) \geq 1\} \) 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 \((\Delta_{\beta}, \Delta_L)\) such that

\[
C_1 \left( M(u)^\perp \right)^2 \geq \left( 1 + c_1(u)^2 \right) \left( M(u)^\perp \right)^2 \geq \left( M(u)^\perp \right)^2 + (M(u))^2 = \mathbb{E} \left| \sum_{j=1}^p X_j \Delta_{\beta,j} \right|^2.
\]

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 \( C_\lambda \)) that does not depend on \( u \) or \((\Delta_{\beta}, \Delta_L)\) such that

\[
\|\Delta_L\|_F^2 = \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F^2 + \|\mathcal{P}_{\Phi(u)}\Delta_{\beta}\|_F^2 < \left( 1 + 4c_2(u)^2 \right) \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F^2 \leq C_2 \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F^2.
\]

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 \( \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F - \kappa_3 \|\Delta_{\beta}\|_F > 0 \),

\[
\left[ c_1(u) M(u)^\perp - \left( \kappa_3 \|\Delta_{\beta}\|_F + \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F \right) \right]^2 + \left[ M(u)^\perp - c_2(u) \left( \kappa_3 \|\Delta_{\beta}\|_F + \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F \right) \right]^2
= \left( 1 + c_1(u)^2 \right) \left( M(u)^\perp \right)^2 + \left( 1 + c_2(u)^2 \right) \left( \kappa_3 \|\Delta_{\beta}\|_F + \|\mathcal{P}_{\Phi(u)}\Delta_{\beta}\|_F \right)^2
\]

\[
- 2(1 + c_1(u) + c_2(u)) M(u)^\perp \left( \kappa_3 \|\Delta_{\beta}\|_F + \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F \right)
\]

\[
\geq \left( 1 - \frac{c_1(u) + c_2(u)}{\sqrt{(c_1(u) + c_2(u))^2 + (1 - c_1(u) c_2(u))^2}} \right) \left[ \left( M(u)^\perp \right)^2 + \left( \kappa_3 \|\Delta_{\beta}\|_F + \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F \right)^2 \right]
\]

\[
\geq \left( 1 - \frac{\sqrt{3r(u)\kappa_1} + 1/(\sqrt{3r(u)\kappa_1})}{\sqrt{\left( \sqrt{3r(u)\kappa_1} + 1/(\sqrt{3r(u)\kappa_1}) \right)^2 + (1 - 1/C_{\Phi X})^2}} \right) \left[ \left( M(u)^\perp \right)^2 + \left( \kappa_3 \|\Delta_{\beta}\|_F + \|\mathcal{P}_{\Phi(u)}\Delta_L\|_F \right)^2 \right]
\]

\[
\geq \left( 1 - \frac{\sqrt{3r(u)\kappa_1} + 1/(\sqrt{3r(u)\kappa_1})}{\sqrt{\left( \sqrt{3r(u)\kappa_1} + 1/(\sqrt{3r(u)\kappa_1}) \right)^2 + (1 - 1/C_{\Phi X})^2}} \right) \left( \frac{1}{C_1} \mathbb{E} \left| \sum_{j=1}^p X_j \Delta_{\beta,j} \right|^2 + \frac{1}{C_2} \|\Delta_L\|_F^2 \right)
\]

\[
\geq \left( \frac{\sigma_{\min}^2}{C_1} \wedge \frac{1}{C_2} \right) \left( 1 - \frac{\sqrt{3r(u)\kappa_1} + 1/(\sqrt{3r(u)\kappa_1})}{\sqrt{\left( \sqrt{3r(u)\kappa_1} + 1/(\sqrt{3r(u)\kappa_1}) \right)^2 + (1 - 1/C_{\Phi X})^2}} \right) \left( NT \|\Delta_{\beta}\|_F^2 + \|\Delta_L\|_F^2 \right)
\]

\[
\geq \left( \frac{\sigma_{\min}^2}{C_1} \wedge \frac{1}{C_2} \right) \left( 1 - \frac{\sqrt{3r\kappa_1} + 1/(\sqrt{3r\kappa_1})}{\sqrt{\left( \sqrt{3r\kappa_1} + 1/(\sqrt{3r\kappa_1}) \right)^2 + (1 - 1/C_{\Phi X})^2}} \right) \left( NT \|\Delta_{\beta}\|_F^2 + \|\Delta_L\|_F^2 \right)
\]
Comparing equations (S.B.13), (S.B.14) and (S.B.25), since $T > T$ that for a bigger again, since $C$ increases as $k$ increases. So $\psi(C) = 1$/(\sqrt{3rC})$ increases. Hence, $\sqrt{3rC} 2$ and real numbers $\lambda, \kappa, \mu$ where $\lambda, \kappa, \mu$ are by construction. Therefore, $\sqrt{3rC}$ increases. So $\sqrt{3rC} 2$ ≤ $\sqrt{3rC}$. Let

$$C_{\text{RSC}} := \left( \frac{\sigma_{\text{min}}}{C_1} \land \frac{1}{C_2} \right) \left( 1 - \frac{\sqrt{3rC} + 1/(\sqrt{3rC})}{\sqrt{\left(\sqrt{3rC} + 1/(\sqrt{3rC})\right)^2 + (1 - 1/C_{\text{X}})^2}} \right).$$

(S.B.24)

Again, since $\sqrt{3rC} > 1$, $C_{\text{RSC}}$ gets bigger when $\kappa$ and $C_2$ get smaller, which is a consequence of a bigger $C_\lambda$. Meanwhile, $C_{\text{RSC}}$ 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,

$$\inf_{u \in \{u : \|x\|_{\text{F}} \geq 1\}} \left( E \left\| \sum_{j=1}^p X_j \Delta_{\beta,j} + \Delta_L \right\|_{\text{F}}^2 - C_{\text{RSC}} \left( NT \|\Delta_{\beta}\|_{\text{F}}^2 + \|\Delta_L\|_{\text{F}}^2 \right) \right) \geq 0.$$

(S.B.25)

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

$$\inf_{u \in \{u : \|x\|_{\text{F}} = 0\}} \left( E \left\| \sum_{j=1}^p X_j \Delta_{\beta,j} + \Delta_L \right\|_{\text{F}}^2 - C_{\text{RSC}} \left( NT \|\Delta_{\beta}\|_{\text{F}}^2 + \|\Delta_L\|_{\text{F}}^2 \right) \right) \geq 0.$$

$$\inf_{u \in \{u : \|x\|_{\text{F}} = 0\}} \left( E \left\| \sum_{j=1}^p X_j \Delta_{\beta,j} + \Delta_L \right\|_{\text{F}}^2 - C_{\text{RSC}} \left( NT \|\Delta_{\beta}\|_{\text{F}}^2 + \|\Delta_L\|_{\text{F}}^2 \right) \right) \geq 0.$$

(S.B.26)

\footnote{To see this, let $\psi := \frac{\|x\|_{\text{F}} + \|y\|_{\text{F}}}{\sqrt{(k_1 + k_2)^2 + (1 - k_1k_2)^2}}$. 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, $(1 + k_1^2)x^2 + (1 + k_2^2)y^2 - 2(k_1 + k_2)xy \geq (1 - \psi)((1 + k_1^2)x^2 + (1 + k_2^2)y^2) \geq (1 - \psi)(x^2 + y^2)$.}
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 \( \varepsilon_{NT} < \frac{C_{RSC}}{2} \). 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 \( \varepsilon_{NT} < \frac{C_{RSC}}{2} \) 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 \( G_u(f(Z_{it})) := \sum_i[f(Z_{it}) - \mathbb{E}(f(Z_{it}))/\sqrt{NT} \). 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 \( \gamma \) defined in equation (3.4),

\[
P \left( \sup_{u \in U} \left| G_u \left( \rho_u \left( V_{it}(u) - X_{it}^\beta - \Delta L_{it} \right) - \rho_u(V_{it}(u)) \right) \right| \leq C_{sup} \sqrt{\log(NT) \left( \sqrt{p \log((p + 1)NT)} \vee \sqrt{f(N \vee T)} \right) \gamma} \right) \to 1.
\]

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

Let \( \Omega_0 \) 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 \( G_u \) in Lemma S.B.4 holds. By Lemmas 1 and S.B.4 and by equation (3.1) under \( \Omega_L \), the event \( \Omega_0 \) occurs w.p.a.1. It is then sufficient to show that the following event
has zero probability under $\Omega_0$:

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

where $\gamma = C_{\text{error}} \alpha^2 NT \left( (1 + C_\lambda) \vee \sqrt{\log(NT)} \right) \left( \sqrt{p \log((p + 1)NT)/NT} \vee \sqrt{r/(N \wedge T)} \right)$ for some $C_{\text{error}} > 0$ 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

$$0 \geq \inf_{\|\Delta_\beta\|^2_F + \frac{1}{NT}\|\Delta_L\|^2_F = \gamma^2} \frac{1}{NT} \mathbb{E} \left[ \rho_u (V(u) - \sum_{j=1}^{p} X_j \Delta_{\beta,j} - \Delta_L) - \rho_u (V(u)) \right]$$

$$+ \lambda \left[ \|L_0(u) + \Delta_L\|_* - \|L_0(u)\|_* \right] \quad (\text{S.B.27})$$

$$= \inf_{\|\Delta_\beta\|^2_F + \frac{1}{NT}\|\Delta_L\|^2_F = \gamma^2} \frac{1}{NT} \mathbb{E} \left[ \rho_u (V(u) - \sum_{j=1}^{p} X_j \Delta_{\beta,j} - \Delta_L) - \rho_u (V(u)) \right]$$

$$+ \frac{1}{\sqrt{NT}} \mathcal{G}_u \left( \rho_u \left( V_{it}(u) - X_{it}' \Delta_{\beta} - \Delta_{L,it} \right) - \rho_u (V_{it}(u)) \right)$$

$$+ \lambda \left[ \|L_0(u) + \Delta_L\|_* - \|L_0(u)\|_* \right]. \quad (\text{S.B.28})$$

For the expectation, Theorem 1 implies that

$$\inf_{u \in \mathcal{U}} \frac{1}{NT} \mathbb{E} \left[ \rho_u (V(u) - \sum_{j=1}^{p} X_j \Delta_{\beta,j} - \Delta_L) - \rho_u (V(u)) \right]$$

$$\geq \frac{C_{\text{min}} C_{\text{RSC}}}{2\alpha^2_{\text{NT}}} \left( \|\Delta_\beta\|^2_F + \frac{\|\Delta_L\|^2_F}{NT} \right)$$

$$= \frac{C_{\text{min}} C_{\text{RSC}}}{2\alpha^2_{\text{NT}}} \gamma^2, \quad (\text{S.B.29})$$

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

For the error process $\mathcal{G}_u/\sqrt{NT}$, by ii) in $\Omega_0$,

$$\frac{1}{\sqrt{NT}} \sup_{u \in \mathcal{U}} \left| \mathcal{G}_u \left( \rho_u \left( V_{it}(u) - X_{it}' \Delta_{\beta} - \Delta_{L,it} \right) - \rho_u (V_{it}(u)) \right) \right|$$
\[ \leq C_{\text{sup}} \sqrt{\log(NT)} \left( \sqrt{\frac{p \log((p+1)NT)}{NT}} \vee \sqrt{\frac{\bar{r}}{N \wedge T}} \right) \gamma. \]  

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

\[
\sup_{u \in \mathcal{U}} \lambda \|L_0(u) + \Delta_L\|_* - \|L_0(u)\|_* \leq \lambda \left( C_{\text{Cone}} \sqrt{p(N \wedge T) \log((p+1)NT)} + \frac{2(1 + C_\lambda)}{C_\lambda} \|P_{\Phi(u)} \Delta_L\|_* \right) 
\]

\[
\leq \lambda \left( C_{\text{Cone}} \sqrt{p(N \wedge T) \log((p+1)NT)} + \frac{2\sqrt{3}(1 + C_\lambda)}{C_\lambda} \|\Delta_L\|_F \right) 
\]

\[
\leq (1 + C_\lambda)C_{\text{op}} \left( C_{\text{Cone}} + \frac{2\sqrt{3}(1 + C_\lambda)}{C_\lambda} \right) \left( \frac{p \log((p+1)NT)}{NT} \vee \sqrt{\frac{\bar{r}}{N \wedge T}} \right) \gamma, 
\]

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

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

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

with probability one. Hence, under \( \Omega_0 \), inequality (S.B.27) and thus inequality (S.B.26) hold with zero probability. Since \( \Omega_0 \) holds w.p.a.1 (implicitly conditional on \( (W_L, \Omega_L) \)), we obtain the desired result.

Remark S.B.3. As noted in Section 3.1, all the arguments are implicitly conditional on \( (W_L, \Omega_L) \). 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,
\[
\max_{k \in \{1, \ldots, N\}} \{|\hat{\sigma}_k(u) - \sigma_k(u)|\} \leq \|\hat{\Delta}_L(u)\| \leq \|\Delta_L(u)\|_F,
\]
with probability one. Hence, \(\sup_{u \in U} \max_k \{|\hat{\sigma}_k(u) - \sigma_k(u)|\} \leq \sup_{u \in U} \|\hat{\Delta}_L(u)\|_F\). 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 \(\{\hat{\sigma}_r(u) \geq C_r\} \cap \{\hat{\sigma}_{r(u)+1} < C_r\}\). The latter event, under the event \(\Omega_{sv2} = \{\sigma_{r(u)}\text{ is of the order of }\sqrt{NT}\}\), can be implied by \(\Omega_{sv1} := \{|\hat{\sigma}_r(u) - \sigma_r(u)| \leq \sqrt{NT\gamma}, |\hat{\sigma}_{r(u)+1} - 0| \leq \sqrt{NT\gamma}\}\) for sufficiently large \(N\) and \(T\) by the choice of \(C_r\). The desired result is thus obtained since w.p.a.1, \(\Omega_{sv1}\) is true by Corollary 1 and \(\Omega_{sv2}\) 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 \(\Omega_1\) be the event that \(\max_{1 \leq j \leq p} \|X_j\|_F^2 \leq C_X NT\). Under Assumption 1, \(\mathbb{P}(\Omega_1) \to 1\). Recall that the \((i, t)\)-th element in subgradient \(\nabla \rho_u(V(u))\) is
\[
(\nabla \rho_u(V(u)))_{it} = u \mathbb{1}(Vi(u) \geq 0) + (u - 1) \mathbb{1}(Vi(u) < 0),
\]
with probability one. By Assumption 1 and by \(V(u) = Y - q_{r|W_X}(u)\), the elements in \(\nabla \rho_u(V(u))\) 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 \(M = 5\sqrt{2C_X NT \log((p + 1)NT)}\). Note that
\[
\mathbb{P}\left(\sup_{u \in U} \sup_{1 \leq j \leq p} \frac{|\langle \nabla \rho_u(V(u), X_j \rangle|}{M} > M \right) \leq \mathbb{P}\left(\sup_{u \in U} \sup_{1 \leq j \leq p} \left|\langle \nabla \rho_u(V(u), X_j \rangle > M \right) \left|\Omega_1 \right\) \mathbb{P}(\Omega_1) + \mathbb{P}(\Omega_2^c).\right)
\]
Since \(\mathbb{P}(\Omega_2^c) \to 0\), it is sufficient to show the conditional probability in equation (S.B.32) converges to zero.

Let \(U_K = (u_1, u_2, ..., u_K)\) be an \(\varepsilon\)-net of \(U\). Let \(\varepsilon = \frac{1}{\sqrt{NT}}\) and \(K \varepsilon \leq 1\). By the triangle
inequality,

$$\sup_{u \in U \atop 1 \leq j \leq p} \left| \langle \nabla \rho_u(V(u)), X_j \rangle \right| \leq \max_{u_k \in U_k} \left| \langle \nabla \rho_{u_k}(V(u_k)), X_j \rangle \right| + \sup_{u \in U \atop 1 \leq j \leq p} \left| \langle \nabla \rho_u(V(u)) - \nabla \rho_{u_k}(V(u_k)), X_j \rangle \right|$$

$$=: \mathcal{E}_1 + \mathcal{E}_2.$$ 

Then by the union bound,

$$\mathbb{P} \left( \sup_{u \in U \atop 1 \leq j \leq p} \left| \langle \nabla \rho_u(V(u)), X_j \rangle \right| > M \bigg| \Omega_1 \right) \leq 2 \max \left\{ \mathbb{P} \left( \mathcal{E}_1 > \frac{1}{5} M \bigg| \Omega_1 \right), \mathbb{P} \left( \mathcal{E}_2 > \frac{4}{5} M \bigg| \Omega_1 \right) \right\}.$$ 

### Bound on $\mathcal{E}_1$.

By $K \leq 1/\varepsilon$, we have

$$\mathbb{P} \left( \mathcal{E}_1 > \frac{1}{5} M \bigg| \Omega_1 \right) \leq \frac{p}{\varepsilon} \max_{u_k \in U_k} \mathbb{P} \left( \left| \langle \nabla \rho_{u_k}(V(u_k)), X_j \rangle \right| \geq \frac{1}{5} M \bigg| \Omega_1 \right)$$

$$= \frac{p}{\varepsilon} \max_{u_k \in U_k} \mathbb{E} \left[ \mathbb{P} \left( \left| \langle \nabla \rho_{u_k}(V(u_k)), X_j \rangle \right| \geq \frac{1}{5} M \bigg| \Omega_1, W_X \right) \bigg| \Omega_1 \right]$$

$$\leq 2p \sqrt{NT} \exp \left( -\frac{M^2}{50C_X NT} \right)$$

$$= \frac{2p}{(p + 1)\sqrt{NT}} \to 0.$$ 

(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 $\Omega_1$ under $\varepsilon = 1/\sqrt{NT}$.

### Bound on $\mathcal{E}_2$.

By definition, the $(i, t)$-th element in $\nabla \rho_u(V(u)) - \nabla \rho_{u_k}(V(u_k))$ is almost surely

$$u \mathbb{1}(V_{it}(u) \geq 0) + (u - 1) \mathbb{1}(V_{it}(u) < 0) - [u_k \mathbb{1}(V_{it}(u_k) \geq 0) + (u_k - 1) \mathbb{1}(V_{it}(u_k) < 0)]$$

$$= (u - u_k) + \mathbb{1}(V_{it}(u_k) < 0) - \mathbb{1}(V_{it}(u) < 0).$$ 

(S.B.35)

Let $\Xi_1$ and $\Xi_2$ be two $N \times T$ matrices whose $(i, t)$-th elements are

$$\Xi_{1, it} := u - u_k$$

(S.B.36)

$$\Xi_{2, it} := \mathbb{1}(V_{it}(u_k) < 0) - \mathbb{1}(V_{it}(u) < 0).$$ 

(S.B.37)
Then by equation (S.B.35), \( \nabla \rho_u(V(u)) - \nabla \rho_{u_k}(V(u_k)) = \Xi_1 + \Xi_2 \). Therefore,

\[
\mathbb{P} \left( \mathcal{E}_2 > \frac{4}{5} M \left| \Omega_1 \right. \right) \leq \mathbb{P} \left( \sup_{|u-u_k| \leq \varepsilon, u_k \in U_K} |\langle \Xi_1, X_j \rangle| > \frac{2}{5} M \left| \Omega_1 \right. \right) + \mathbb{P} \left( \sup_{|u-u_k| \leq \varepsilon, u_k \in U_K} |\langle \Xi_2, X_j \rangle| > \frac{2}{5} M \left| \Omega_1 \right. \right)
\]

\[
=: \mathbb{P}_1 + \mathbb{P}_2.
\]

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

\[
\mathbb{P} \left( \sup_{1 \leq j \leq p} \left| \langle \Xi_1, X_j \rangle \right| > \frac{2}{5} M \left| \Omega_1 \right. \right) \leq \mathbb{P} \left( \sup_{1 \leq j \leq p} \|\Xi_1\|_F \|X_j\|_F > \frac{2}{5} M \left| \Omega_1 \right. \right)
\]

\[
\leq \mathbb{P} \left( \varepsilon \sqrt{C_X N T} > \frac{2}{5} M \right)
\]

\[
= \mathbb{P} \left( \sqrt{C_X N T} > 2 \sqrt{2C_X N T \log((p+1)NT)} \right)
\]

\[
= 0
\]

(S.B.38)

for large enough \( N \) and \( T \). The second inequality is by the definition of \( \Xi_1 \) and \( \Omega_1 \). 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

\[
\Xi_{2, it}^{(1)}(u_k) := \mathbb{I}(V_{it}(u_k) < 0) - \mathbb{I}(V_{it}(u_k - \varepsilon) < 0).
\]

(S.B.40)

\[
\Xi_{2, it}^{(2)}(u_k) := \mathbb{I}(V_{it}(u_k) < 0) - \mathbb{I}(V_{it}(u_k + \varepsilon) < 0).
\]

(S.B.41)

Consider an arbitrary element in \( \Xi_2 \): \( \Xi_{2, it} := \mathbb{I}(V_{it}(u_k) < 0) - \mathbb{I}(V_{it}(u) < 0) \). By Assumption 1, \( V_{it}(u) \) is strictly decreasing in \( u \) almost surely. Hence, \( \mathbb{I}(V_{it}(u) < 0) \) 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.

\[
\sup_{|u-u_k| \leq \varepsilon, u_k \in U_K} |\langle \Xi_2, X_j \rangle| \leq \sup_{u_k-\varepsilon \leq u \leq u_k, u_k \in U_k} |\langle \Xi_2, X_j \rangle| + \sup_{u_k \leq u \leq u_k + \varepsilon, u_k \in U_k} |\langle \Xi_2, X_j \rangle|
\]

\[
\leq \sup_{u_k-\varepsilon \leq u \leq u_k, u_k \in U_k} |\langle \Xi_2, |X_j| \rangle| + \sup_{u_k \leq u \leq u_k + \varepsilon, u_k \in U_k} |\langle \Xi_2, |X_j| \rangle|
\]

\[
\leq \max_{u_k \in U_K} |\langle \Xi_{2}^{(1)}, |X_j| \rangle| + \max_{u_k \in U_K} |\langle \Xi_{2}^{(2)}, |X_j| \rangle|.
\]

21
The first inequality is elementary. To see why the second inequality holds, note that the elements in $\bar{\Xi}_2$ 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 $\Xi_2$ 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 $X_j$’s so that they also have the same sign. The third inequality then follows because now that elements in both $|X_j|$ and in $\Xi_2$ have the same signs, the two absolute inner products in the second line increase as the magnitude of any of the elements in $\Xi_2$ increases. Therefore,

$$P_2 \leq P\left(\max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |\langle\Xi^{(1)}_2, X_j\rangle| > \frac{1}{5} M \bigg| \Omega_1\right) + P\left(\max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |\langle\Xi^{(2)}_2, X_j\rangle| > \frac{1}{5} M \bigg| \Omega_1\right). \tag{S.B.42}$$

Let us first bound $\max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |\langle\Xi^{(1)}_2, X_j\rangle|$. The expectation of an arbitrary element $\Xi^{(1)}_{2,it}$ in $\Xi^{(1)}_2$ satisfies

$$\mathbb{E}\left(\Xi^{(1)}_{2,it}(u_k) \big| W_X\right) = P\left(V_{it}(u_k) < 0 \leq V_{it}(u_k - \varepsilon) \big| W_X\right) = P\left(q_{Y_{it}|W_X}(u_k - \varepsilon) \leq Y_{it} < q_{Y_{it}|W_X}(u_k) \big| W_X\right) = \varepsilon$$

where the second equality is by the definition of $V_{it}(u)$. Let $\tilde{\Xi}^{(1)}_2 = \mathbb{E}\left(\Xi^{(1)}_2(u_k) \big| W_X\right)$ be an $N \times T$ matrix whose elements are all equal to $\varepsilon$. Under $\Omega_1$, by the Cauchy-Schwarz inequality and by $\varepsilon = 1/\sqrt{NT}$, we have $\max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |\langle\tilde{\Xi}^{(1)}_2, X_j\rangle| \leq \varepsilon \sqrt{C_X NT} = \sqrt{C_X NT}$ with probability one. Therefore,

$$P\left(\max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |\langle\Xi^{(1)}_2, X_j\rangle| > \frac{1}{5} M \bigg| \Omega_1\right) \leq P\left(\max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |\langle\Xi^{(1)}_2 - \tilde{\Xi}^{(1)}_2, X_j\rangle| + \max_{1 \leq j \leq p} |\langle\tilde{\Xi}^{(1)}_2, X_j\rangle| > \frac{1}{5} M \bigg| \Omega_1\right) \leq P\left(\max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |\langle\Xi^{(1)}_2 - \tilde{\Xi}^{(1)}_2, X_j\rangle| > \frac{1}{5} M - \sqrt{C_X NT} \bigg| \Omega_1\right) \leq \frac{P}{\varepsilon} \max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} \mathbb{E}\left(\left|\langle\Xi^{(1)}_2 - \tilde{\Xi}^{(1)}_2, X_j\rangle\right| > \frac{1}{5} M - \sqrt{C_X NT} \bigg| W_X\right) \bigg| \Omega_1\right) \leq 2p\sqrt{NT} \exp\left(-\frac{(M/5 - \sqrt{C_X NT})^2}{2C_X NT}\right) \to 0. \tag{S.B.43}$$

The penultimate equality is by the law of iterated expectation since $\max_{1 \leq j \leq p} \|X_j\|^2_F$ in $\Omega_1$ is
By the triangle inequality, we have
\[ W \]
Here \( C \) is an absolute constant and \( \epsilon \) is an absolute constant.

Finally, we can show that \( \mathbb{P} \left( \max_{u_k \in \mathcal{U}_k, 1 \leq j \leq p} |(\bar{X}_2^{(1)} - \bar{X}_2^{(2)})| > M/5|\Omega_1 \right) \rightarrow 0 \) 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 \( \|A\| \leq CQ \left( \sqrt{N} + \sqrt{T} + t \right) \) with probability at least \( 1 - 2\exp(-t^2) \).

Here \( C \) is an absolute constant and \( Q = \max_{i,j} \|A_{ij}\|_{\psi_2} \).

For any \( u \), entries in matrix \( \nabla \rho_u(V(u)) \) are independent mean zero random variables conditional on \( W_X \). Since each of them is Bernoulli, they are sub-Gaussian with \( \|\nabla \rho_u(V(u))\|_{\psi_2} = \inf \{ c > 0 : (1 - u) \exp(u^2/c^2) + u \exp((1 - u^2)/c^2) \leq 2 \} \). Take \( Q = \sup_{u \in (0,1)} \|\nabla \rho_u(V(u))\|_{\psi_2} < \infty \). 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 \( \varepsilon \)-net of \( \mathcal{U} \) with \( \varepsilon K \leq 1 \). This time let \( \varepsilon = 1/\sqrt{N \vee T} \).

By the triangle inequality, we have
\[
\sup_{u \in \mathcal{U}} \|\nabla \rho_u(V(u))\| \leq \max_{u_k \in \mathcal{U}_K} \|\nabla \rho_{u_k}(V(u_k))\| + \sup_{|u - u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} \|\nabla \rho_u(V(u)) - \nabla \rho_{u_k}(V(u_k))\|
\]
\[ =: \mathcal{F}_1 + \mathcal{F}_2. \quad \text{(S.B.44)} \]

By equation (S.B.44),
\[
\mathbb{P} \left( \sup_{u \in \mathcal{U}} \|\nabla \rho_u(V(u))\| > M \right) \leq 2 \max \{ \mathbb{P} (\mathcal{F}_1 > M/2), \mathbb{P} (\mathcal{F}_2 > M/2) \}. \quad \text{(S.B.45)}
\]

**Bound on \( \mathcal{F}_1 \).** By \( \varepsilon K \leq 1 \), we have
\[
\mathbb{P} (\mathcal{F}_1 > M/2) \leq \frac{1}{\varepsilon} \max_{u_k \in \mathcal{U}_K} \mathbb{P} (\|\nabla \rho_{u_k}(V(u_k))\| > M/2)
\]
\[
= \frac{1}{\varepsilon} \max_{u_k \in \mathcal{U}_K} \mathbb{E} [\mathbb{P} (\|\nabla \rho_{u_k}(V(u_k))\| > M/2|W_X)]
\]
\[
\leq 2\sqrt{N \vee T} \exp(- (N \vee T)) \rightarrow 0 \quad \text{(S.B.46)}
\]

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,

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

$$
\sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} \|\nabla \rho_u(V(u)) - \nabla \rho_{u_k}(V(u_k))\| \leq \sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} \|\Xi_1\| + \sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} \|\Xi_2\|
$$

$$
\leq \varepsilon \sqrt{NT} + \sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} \|\Xi_2\|
$$

$$
= \sqrt{N \vee T} + \sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} \|\Xi_2\|
$$

where $\Xi_1$ and $\Xi_2$ 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 $\Xi_1$ are equal to $u - u_k$, whose magnitude is bounded by $\varepsilon$ 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,

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

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,

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

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 $\Xi_2$ are all nonnegative when $u < u_k$ and all nonpositive when $u > u_k$, and in either case, we have $|\Xi_{2, it}| \leq |\Xi_{2, it}^{(l)}|$, $l = 1, 2$ for all $i, t$. Inequality (S.B.49) is thus implied by
inequality (S.B.48).

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

$$\mathbb{P} \left( \max_{u_k \in U_k} \|\Xi_2^{(1)}\| > M/4 - \sqrt{N \wedge T}/2 \right) \leq \mathbb{P} \left( \max_{u_k \in U_k} \|\Xi_2^{(1)}\| + \max_{u_k \in U_k} \|\tilde{\Xi}_2^{(1)}\| > C_{op} \sqrt{N \vee T}/4 - \sqrt{N \wedge T}/2 \right) \leq \mathbb{P} \left( \max_{u_k \in U_k} \|\Xi_2^{(1)} - \tilde{\Xi}_2^{(1)}\| > C_{op} \sqrt{N \vee T}/4 - 3\sqrt{N \wedge T}/2 \right) \leq K \max_{u_k \in U_k} \mathbb{P} \left( \|\Xi_2^{(1)} - \tilde{\Xi}_2^{(1)}\| > (C_{op}/4 - 3/2) \sqrt{N \vee T} \right) \leq \sqrt{N \vee T} \max_{u_k \in U_k} \mathbb{E} \left[ \mathbb{P} \left( \|\Xi_2^{(1)} - \tilde{\Xi}_2^{(1)}\| > (C_{op}/4 - 3/2) \sqrt{N \vee T} \mid W_X \right) \right] \leq \frac{2\sqrt{N \vee T}}{\exp(N \vee T)} \to 0,$$  

(S.B.50)

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

Similarly, we have $\mathbb{P} \left( \max_{u_k \in U_k} \|\Xi_2^{(2)}\| > M/4 - \sqrt{N \wedge T}/2 \right) \to 0$ 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 $\Delta_L$. Its proof is in this section.

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

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

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

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

25
Let \( w_1 = V_{it}(u) \) and \( w_2 = X'_{it}\Delta + \Delta_{L,it} \) where \( \Delta \) and \( \Delta_L \) 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,

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

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

\[
\mathbb{E} \left( \int_0^{X'_{it}\Delta + \Delta_{L,it}} \left( 1(0 \leq s \leq \Delta_{L,it}) - 1(0 \leq \Delta_{L,it} \leq 0) \right) ds \right) \\
\geq \mathbb{E} \left( \int_0^{\kappa_{it} \cdot (X'_{it}\Delta + \Delta_{L,it}) (1 \land \delta)} \left( 1(0 \leq s \leq \Delta_{L,it}) - 1(0 \leq \Delta_{L,it} \leq 0) \right) ds \right) \\
= \mathbb{E} \left[ \mathbb{E} \left( \int_0^{\kappa_{it} \cdot (X'_{it}\Delta + \Delta_{L,it}) (1 \land \delta)} \left( 1(0 \leq s \leq \Delta_{L,it}) - 1(0 \leq \Delta_{L,it} \leq 0) \right) ds \right) | W_X \right] \\
= \mathbb{E} \int_0^{\kappa_{it} \cdot (X'_{it}\Delta + \Delta_{L,it}) (1 \land \delta)} \left( F_{V_{it}(u)\mid W_X}(s) - F_{V_{it}(u\mid W_X}(0) \right) ds \\
= \mathbb{E} \int_0^{\kappa_{it} \cdot (X'_{it}\Delta + \Delta_{L,it}) (1 \land \delta)} s f_{V_{it}(u)\mid W_X}(\tilde{s}(s)) ds \\
\geq \mathbb{E} \left[ \frac{\kappa_{it}^2 (1 \land \delta)^2 (X'_{it}\Delta + \Delta_{L,it})^2}{2} \right] \\
= \frac{(1 \land \delta)^2}{2} \cdot \mathbb{E} \left[ \frac{(X'_{it}\Delta + \Delta_{L,it})^2}{2} \right] \tag{S.B.52}
\]

where \( f \) 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 \( \kappa_{it} \cdot (X'_{it}\Delta + \Delta_{L,it}) (1 \land \delta) \in [-\delta,\delta] \). The last equality is by the definition of \( \kappa_{it} \). Now we lower bound the expectation on the right hand side of the last equality. We have

\[
\mathbb{E} \left[ \frac{(X'_{it}\Delta + \Delta_{L,it})^2}{2\alpha_{NT} + \sqrt{X'_{it}X_{it}}\|\Delta\|_F \lor 1} \right] \\
\geq \mathbb{E} \left[ \frac{1}{2} \left( X'_{it}X_{it} \leq \alpha_{NT}/\gamma \right) (X'_{it}\Delta + \Delta_{L,it})^2 \right] \\
\frac{\sqrt{X_{it}X_{it}}\|\Delta\|_F \lor 1}{2\alpha_{NT} + \sqrt{X'_{it}X_{it}}\|\Delta\|_F \lor 1}
\]
\[
\begin{align*}
&\geq \mathbb{E} \left[ \mathbb{I} \left( \sqrt{X_{it}'X_{it}} \leq \alpha_{NT}/\gamma \right) \right] (X_{it}'\Delta_{\beta} + \Delta_{L,it})^2 \\
&\quad \quad \quad \div (2\alpha_{NT} + (\alpha_{NT}\|\Delta_{\beta}\|_F/\gamma \vee 1))^2 \\
&\geq \frac{\mathbb{E} \left[ (X_{it}'\Delta_{\beta} + \Delta_{L,it})^2 \right]}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_{\beta}\|_F/\gamma \vee 1))^2} - \mathbb{E} \left[ \mathbb{I} \left( \sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma \right) \right] \cdot (X_{it}'\Delta_{\beta} + \Delta_{L,it})^2 \\
&\quad \quad \quad \div (2\alpha_{NT} + (\alpha_{NT}\|\Delta_{\beta}\|_F/\gamma \vee 1))^2 \\
&\geq \frac{\mathbb{E} \left[ (X_{it}'\Delta_{\beta} + \Delta_{L,it})^2 \right] - \varepsilon_{NT} (\|\Delta_{\beta}\|^2_F + \Delta_{L,it}^2)}{(2\alpha_{NT} + (\alpha_{NT}\|\Delta_{\beta}\|_F/\gamma \vee 1))^2} \\
&\quad \quad \quad \quad \quad \text{(S.B.53)}
\end{align*}
\]

where the third inequality is by Cauchy-Schwarz. The last inequality is by substituting \(\varepsilon_{NT} := 2 \left( \max_{i,t} \mathbb{E} \left[ \mathbb{I} \left( \sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma \right) \cdot X_{it}'X_{it} \right] \vee \max_{i,t} \mathbb{P} \left( \sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma \right) \right) \) 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), \(\gamma = O(\alpha_{NT}^2\sqrt{\log(NT)/N \wedge T})\) because \(C_\lambda = O \left( \sqrt{\log(NT)} \right) \). Therefore, \(\sqrt{p} = o(\alpha_{NT}/\gamma)\).

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

\[
\max_{i,t} \mathbb{P} \left( \sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma \right) \leq p \max_{j=1,\ldots,p;i,t} \mathbb{P} \left( |X_{j,it}| > \frac{\alpha_{NT}}{\sqrt{p+1} \gamma} \right) \leq p \max_{j=1,\ldots,p;i,t} \mathbb{E} \left( \varphi(\frac{|X_{j,it}|}{\varepsilon}) \right) \to 0
\]

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 \(\max_{j=1,\ldots,p;i,t} \mathbb{E} \left( \varphi(\frac{|X_{j,it}|}{\varepsilon}) \right) < C_\varphi\) by Assumption 3 ii).

Now we show that \(\max_{i,t} \mathbb{E} \left( \mathbb{I} \left( \sqrt{X_{it}'X_{it}} > \alpha_{NT}/\gamma \right) \cdot X_{it}'X_{it} \right)\) converges to zero as well.
We have
\[
\begin{align*}
\max_{i,t} \mathbb{E} & \left( \mathbb{1} \left( \sqrt{X'_{it}X_{it}} > \alpha_{NT}/\gamma \right) \cdot X'_{it}X_{it} \right) \\
\leq & \max_{i,t} \left[ \mathbb{E}[X'_{it}X_{it}]^{1+\eta/2} \right] ^{2/(2+\eta)} \cdot \left( \mathbb{P} \left( \sqrt{X'_{it}X_{it}} > \alpha_{NT}/\gamma \right) \right)^{\eta/(2+\eta)} \\
\leq & \max_{j=1,\ldots,p;i,t} \left( \mathbb{E}[|X_{j,it}|]^{2+\eta} \right) ^{2/(2+\eta)} \left( \max_{i,t} \mathbb{P} \left( \sqrt{X'_{it}X_{it}} > \alpha_{NT}/\gamma \right) \right)^{\eta/(2+\eta)} \\
\leq & \left( \max_{j=1,\ldots,p;i,t} \mathbb{E}[\varphi(|X_{j,it}|)] \right) P \cdot \left( \frac{p}{\varphi(\alpha_{NT}/\sqrt{p+1})} \right)^{\eta/(2+\eta)} \\
\to & 0
\end{align*}
\]
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 \( \gamma \).

Proof of Lemma S.B.3

Let \( A(u) := C_{\Phi_X} \kappa_3 \| \Delta_{\beta} \|_F - \varepsilon_0 M(u) \) where \( \varepsilon_0 \) is the same as in Assumption 4 i). We have
\[
\inf_{u \in \mathcal{U}} \left[ M(u) - C_{\Phi_X} \sqrt{3r(u)} \kappa_1 \left( M(u) + \kappa_3 \| \Delta_{\beta} \|_F \right) \right] \\
= \min \left\{ \inf_{u \in \{ u \in \mathcal{U} : A(u) < 0 \}} \left[ M(u) - C_{\Phi_X} \sqrt{3r(u)} \kappa_1 \left( M(u) + \kappa_3 \| \Delta_{\beta} \|_F \right) \right] , \right. \\
\left. \inf_{u \in \{ u \in \mathcal{U} : A(u) \geq 0 \}} \left[ M(u) - C_{\Phi_X} \sqrt{3r(u)} \kappa_1 \left( M(u) + \kappa_3 \| \Delta_{\beta} \|_F \right) \right] \right\}.
\]
\[(S.B.54)\]

First consider the case when \( A(u) < 0 \). The following inequalities hold:
\[
\begin{align*}
\inf_{u \in \{ u \in \mathcal{U} : A(u) < 0 \}} & \left[ M(u) - C_{\Phi_X} \sqrt{3r(u)} \kappa_1 \left( M(u) + \kappa_3 \| \Delta_{\beta} \|_F \right) \right] \\
\geq & \inf_{u \in \{ u \in \mathcal{U} : A(u) < 0 \}} \left( M(u) - (C_{\Phi_X} + \varepsilon_0) \sqrt{3r(u)} \kappa_1 M(u) \right) \\
\geq & \inf_{u \in \mathcal{U}} \left( M(u) - (C_{\Phi_X} + \varepsilon_0) \sqrt{3r(u)} \kappa_1 M(u) \right) \\
\geq & 0,
\end{align*}
\]
\[(S.B.55)\]
where the last inequality is by Assumption 4 i).

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

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

where the inequality is by equation (S.B.10). Meanwhile, recall that the order of $\kappa_3$ is the same as $\kappa_2$ which is $o(N \wedge T)$ uniformly in $u$. So $M(u)^\perp$ must be at least $\sqrt{NT} \| \Delta_\beta \|_F$ uniformly in $u \in \{ u \in U : A(u) \geq 0 \}$, 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 $T > T_3$,

$$
\inf_{u \in \{ u \in U : A(u) \geq 0 \}} \left[ M(u)^\perp - C_{\Phi,X} \sqrt{3r(u)} \kappa_1 \left( M(u) + \kappa_3 \| \Delta_\beta \|_F \right) \right]
\geq \inf_{u \in \{ u \in U : A(u) \geq 0 \}} \left[ M(u)^\perp - \kappa_1 \sqrt{3r(u)} C_{\Phi,X} \left( \frac{C_{\Phi,X}}{\varepsilon_0} + 1 \right) \kappa_3 \| \Delta_\beta \|_F \right]
\geq 0.
$$

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 $\Delta_L$.

Let

$$
\mathcal{A}(\gamma) := \sup_{\{u \in \mathcal{U} : A(u) \geq 0\}} \mathbb{E} \left| \mathcal{G}_u \left( \rho_u (V_{it}(u) - X_{it}' \Delta_\beta - \Delta_{L,it}) - \rho_u (V_{it}(u)) \right) \right|.
$$

For arbitrary random variables $Z_{it}$s and a function $f$, denote the symmetrized version of $\mathcal{G}(f(Z_{it}))$ 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 $\Delta_\beta$ and $\Delta_L$ 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:

$$
\text{Var} \left( \mathcal{G}_u (\rho_u (V_{it}(u) - X_{it}' \Delta_\beta - \Delta_{L,it}) - \rho_u (V_{it}(u))) \right)
$$
\[
\frac{1}{NT} \mathbb{E} \left( \left\| \sum_{j=1}^{p} X_j \delta_{j,i} + \Delta_L \right\|_F^2 \right) \\
\leq \frac{2p}{NT} \sum_{i,t,j=1,\ldots,p} \max_{i,j} \mathbb{E} (X_{j,it}^2) \|\delta_{j,i}\|_F^2 + \frac{2}{NT} \|\Delta_L\|_F^2 \\
\leq 2p \left( \max_{i,t,j=1,\ldots,p} \mathbb{E} (X_{j,it}^2) \right) \|\delta_{j,i}\|_F^2 + \frac{2}{NT} \|\Delta_L\|_F^2 \\
\leq 2 \left( p \max_{i,t,j=1,\ldots,p} \mathbb{E} (\varphi(X_{j,it})) \right) \left( \|\delta_{j,i}\|_F^2 + \frac{1}{NT} \|\Delta_L\|_F^2 \right) \\
\leq 2 (pC_\varphi \lor 1) \gamma^2,
\] (S.B.58)

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

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

Since \(\mathbb{E}(\mathbb{G}_u (\rho_u (V_{it}(u)) - X'_{it} \Delta_\beta - \Delta_{L, it}) - \rho_u (V_{it}(u))) = 0\) 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):

\[\mathbb{P}(A(\gamma) > s) \leq \frac{2}{1 - 8 (pC_\varphi \lor 1) \gamma^2 / s^2} \leq 4 \mathbb{P}(A^0(\gamma) > \frac{s}{4})\]
(S.B.59)

where \(A^0(\gamma)\) is the symmetrized version of \(A(\gamma)\) 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 \(\gamma \sqrt{(pC_\varphi \lor 1) / s} \rightarrow 0\) by the definition of \(s\) for any fixed \(C_{sup}\).

Let \(\Omega_1\) 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

\[\mathbb{P} \left( A^0(\gamma) > \frac{s}{4} \right) \leq \mathbb{P} \left( A^0(\gamma) > \frac{s}{4} \mid \Omega_1 \right) + \mathbb{P}(\Omega_1^c).
\]
(S.B.60)

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

Consider the random variable \(\rho_u (V_{it}(u)) - X'_{it} \Delta_\beta - \Delta_{L, it}) - \rho_u (V_{it}(u))\):

\[\rho_u (V_{it}(u)) - X'_{it} \Delta_\beta - \Delta_{L, it}) - \rho_u (V_{it}(u)) = -u \cdot (X'_{it} \Delta_\beta + \Delta_{L, it}) + \delta_{it} (X'_{it} \Delta_\beta + \Delta_{L, it}, u)\]
where $\delta_t (X'_t \Delta_\beta + \Delta_{L,t}, u) = (V_t(u) - X'_t \Delta_\beta - \Delta_{L,t}) - (V_t(u))_\cdot$. Let

$$B^0_1(\gamma) := \sup_{\|\Delta_\beta\|_F^2 \leq \gamma^2} \left| G^0 (X'_t \Delta_\beta) \right|,$$

$$B^0_2(\gamma) := \sup_{u \in U, (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \atop \|\Delta_\beta\|_F^2 + \frac{1}{N^2} \|\Delta_L\|_F^2 \leq \gamma^2} \left| G^0 (\Delta_{L,t}) \right|,$$

and

$$C^0(\gamma) := \sup_{u \in U, (\Delta_\beta, \Delta_L) \in \mathcal{R}_u \atop \|\Delta_\beta\|_F^2 + \frac{1}{N^2} \|\Delta_L\|_F^2 \leq \gamma^2} \left| G^0 (\delta_t (X'_t \Delta_\beta + \Delta_{L,t}, u)) \right|,$$

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

$$\mathbb{P} \left( A^0(\gamma) \geq \frac{s}{4} \bigg| \Omega_1 \right)$$

$$\leq \mathbb{P} \left( B^0_1(\gamma) + B^0_2(\gamma) + C^0(\gamma) \geq \frac{s}{4} \bigg| \Omega_1 \right)$$

$$\leq 3 \max \left\{ \mathbb{P} \left( B^0_1(\gamma) \geq \frac{s}{12} \bigg| \Omega_1 \right), \mathbb{P} \left( B^0_2(\gamma) \geq \frac{s}{12} \bigg| \Omega_1 \right), \mathbb{P} \left( C^0(\gamma) \geq \frac{s}{12} \bigg| \Omega_1 \right) \right\}. \quad \text{(S.B.61)}$$

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

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

$$\mathbb{P} \left( B^0_1(\gamma) > K_1 \bigg| W_X, \Omega_1 \right)$$

$$\leq \min_{\tau \geq 0} e^{-\tau K_1} \mathbb{E} \left[ \exp (\tau B^0_1(\gamma)) \bigg| W_X, \Omega_1 \right]$$

$$\leq \min_{\tau \geq 0} e^{-\tau K_1} \mathbb{E} \left[ \exp \left( \tau \sup_{\|\Delta_\beta\|_F^2 \leq \gamma^2} \|\Delta_\beta\|_1 \cdot \max_{1 \leq j \leq p} \|G^0 (X_{j,t})\| \right) \bigg| W_X, \Omega_1 \right]$$

$$\leq 2p \min_{\tau \geq 0} e^{-\tau K_1} \max_{1 \leq j \leq p} \mathbb{E} \left[ \exp \left( \tau \sqrt{p} \gamma \cdot G^0 (X_{j,t}) \right) \bigg| W_X, \Omega_1 \right]$$

$$\leq 2p \min_{\tau \geq 0} e^{-\tau K_1} \exp \left( \frac{\tau^2 p \gamma^2 C_X}{2} \right)$$

where the third inequality follows from the fact that $\mathbb{E} [\max_{1 \leq j \leq p} \exp(|z_j|)] \leq 2p \max_{1 \leq j \leq p} \mathbb{E} [\exp(z_j)]$ 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 $\Omega_1$. 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 $\tau$ be any constant and the bound is equal to
\[ P\left( B_1^0(\gamma) > K_1 \mid W_X, \Omega_1 \right) \leq 2p \exp \left( -\frac{K_1^2}{2p\gamma^2C_X} \right) = \frac{2p}{(p+1)NT} \to 0. \]

Therefore,

\[ P\left( B_1^0 > K_1 \mid \Omega_1 \right) = \mathbb{E} \left[ P\left( B_1^0 > K_1 \mid W_X, \Omega_1 \right) \mid \Omega_1 \right] \leq \mathbb{E} \left( \frac{2}{NT} \right) \to 0. \quad (S.B.62) \]

**Bound on** \( B_2^0(\gamma) \). Recall that \( \{\varepsilon_{it}\}_{i,t} \) is the Rademacher sequence in the symmetrized process. Let \( \varepsilon \) be the \( N \times T \) matrix \( (\varepsilon_{it})_{i,t} \). Then with probability one, by defining \( C_{\text{Cone}} := 5\sqrt{2C_X}/(C\lambda C_{\text{op}}) \), we have

\[ B_2^0(\gamma) = \frac{1}{\sqrt{NT}} \sup_{\varepsilon \in \mathbb{R}^N} \left| \sum_{i,t} \varepsilon_{it} \Delta_{L,it} \right| \leq \frac{1}{\sqrt{NT}} \sup_{\varepsilon \in \mathbb{R}^N} \left| \langle \varepsilon, \Delta_L \rangle \right| \]

\[ \leq \frac{1}{\sqrt{NT}} \left\| \varepsilon \right\| \cdot \sup_{(\Delta_{\beta,\Delta_L}) \in \mathbb{R}^N} \left( \left\| \Delta_{\beta} \right\|_2 + \frac{1}{\sqrt{NT}} \left\| \Delta_L \right\|_2 \leq \gamma^2 \right) \]

\[ \leq \frac{1}{\sqrt{NT}} \left\| \varepsilon \right\| \cdot \sup_{\left\| \Delta_{\beta} \right\|_2 + \frac{1}{\sqrt{NT}} \left\| \Delta_L \right\|_2 \leq \gamma^2} \left( C_{\text{Cone}} \sqrt{pN} \log((p+1)NT) \left\| \Delta_{\beta} \right\|_F + \frac{2(1 + C\lambda)}{C\lambda} \left\| P \Phi(u) \Delta_L \right\|_* \right) \]

\[ \leq \frac{1}{\sqrt{NT}} \left\| \varepsilon \right\| \cdot \sup_{\left\| \Delta_{\beta} \right\|_2 + \frac{1}{\sqrt{NT}} \left\| \Delta_L \right\|_2 \leq \gamma^2} \left( C_{\text{Cone}} \sqrt{p(N \wedge T)} \log((p+1)NT) \left\| \Delta_{\beta} \right\|_F + \frac{2\sqrt{3\gamma^2}(1 + C\lambda)}{C\lambda} \left\| \Delta_L \right\|_F \right) \]

\[ \leq \frac{2}{\sqrt{NT}} \left\| \varepsilon \right\| \cdot \left( \left( C_{\text{Cone}} \sqrt{p \log((p+1)NT)} \right) \vee \left( \frac{2(1 + C\lambda)}{C\lambda} \sqrt{3\gamma^2/NT} \right) \right) \gamma, \]

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

\[ K_2 := C_{Sp} \left( \left( C_{\text{Cone}} \sqrt{p \log((p+1)NT)} \right) \vee \left( \frac{2(1 + C\lambda)}{C\lambda} \sqrt{3\gamma^2/NT} \right) \right) \gamma, \]

32
we have
\[ \mathbb{P} \left( \mathcal{B}_2^0(\gamma) > K_2 \Omega_1 \right) \leq \mathbb{P} \left( \| \varepsilon \| > C_{S_p} \sqrt{N \vee T} / 2 \right) \to 0. \] (S.B.63)

**Bound on \( C^0(\gamma) \).** By \( \gamma = o(1) \), it is smaller than one for sufficiently large \( N \) and \( T \). Then let \( \mathcal{U} = \{ u_1, \ldots, u_l \} \) be an \( \epsilon \)-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
\[
\delta_{it}(X'_{it} \Delta_{\beta} + \Delta_{L,it}, u) = \delta_{it}[X'_{it}(\Delta_{\beta} + \beta_0(u) - \beta_0(\bar{u})) + \Delta_{L,it} + L_{0,it}(u) - L_{0,it}(\bar{u}), \bar{u}] \\
- \delta_{it}[X'_{it}(\beta_0(u) - \beta_0(\bar{u})), L_{0,it}(u) - L_{0,it}(\bar{u}), \bar{u}].
\]
Then by the triangle inequality, with probability one we have
\[
C^0(\gamma) \leq \sup_{u \in \mathcal{U}, |u - \bar{u}| < \epsilon, \bar{u} \in \mathcal{U}} \| \mathbb{G}^0(\delta_{it}[X'_{it}(\Delta_{\beta} + \beta_0(u) - \beta_0(\bar{u})) + \Delta_{L,it} + L_{0,it}(u) - L_{0,it}(\bar{u})], \bar{u}] \| \\
+ \sup_{u \in \mathcal{U}, |u - \bar{u}| < \epsilon, \bar{u} \in \mathcal{U}} \| \mathbb{G}^0(\delta_{it}[X'_{it}(\beta_0(u) - \beta_0(\bar{u})), L_{0,it}(u) - L_{0,it}(\bar{u})], \bar{u}] \|. \] (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 \( \Delta_{\beta}s \), and \( \Delta_{L} + L_{0}(u) - L_{0}(\bar{u}) \) and \( L_{0}(u) - L_{0}(\bar{u}) \) as new \( \Delta_{L}s \). 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 \( \Delta_{\beta} \), by Assumption 5, by \( \epsilon = \gamma / \sqrt{p + 1} \) and by \( \| \Delta_{\beta} \|_F \leq \gamma \), we have \( \| \Delta_{\beta} + \beta_0(u) - \beta_0(\bar{u}) \|_F^2 \leq 2(1 + \zeta_\lambda^2/(p + 1)) \gamma^2 \) and \( \| \beta_0(u) - \beta_0(\bar{u}) \|_F^2 \leq \zeta_\lambda^2 \gamma^2 / (p + 1) \) for all \( |u - \bar{u}| \leq \epsilon \). Similarly, for \( \| \Delta_{L} \|_F \leq \sqrt{NT} \gamma \), under equation (3.2), \( \| \Delta_{L} + L_{0}(u) - L_{0}(\bar{u}) \|_F^2 / NT \leq 2(1 + \zeta_\lambda^2/(p + 1)) \gamma^2 \) while \( \| L_{0}(u) - L_{0}(\bar{u}) \|_F^2 / NT \leq \zeta_\lambda^2 \gamma^2 / (p + 1) \) for all \( |u - \bar{u}| \leq \epsilon \). Therefore, we need to expand the ball to be \( \| \Delta_{\beta} \|_F^2 + \| \Delta_{L} \|_F^2 / NT \leq 2(1 + (\zeta_\lambda^2 + \zeta_\lambda^2)/(p + 1)) \gamma^2 \). For simplicity, let \( C_\zeta := 1 + (\zeta_\lambda^2 + \zeta_\lambda^2)/(1 + p) \). Since \( \zeta_\lambda = O(\sqrt{p}) \), \( C_\zeta = O(1) \).

Next, let us expand \( \mathcal{R}_u \). Since rank \( (L_{0}(u) - L_{0}(\bar{u})) \leq r(u) + r(\bar{u}) \leq 2r \), we have \( \| L_{0}(u) - L_{0}(\bar{u}) \|_* \leq \sqrt{2r} \| L_{0}(u) - L_{0}(\bar{u}) \|_F \leq \sqrt{2r} \zeta_\lambda \sqrt{NT} \gamma / \sqrt{p + 1} \) by \( \epsilon = \gamma / \sqrt{p + 1} \) for all \( |u - \bar{u}| \leq \epsilon \). Similarly, for \( \| \Delta_{\beta} + \Delta_{L} \|_F \leq \sqrt{2r} \zeta_\lambda \sqrt{NT} \gamma / \sqrt{p + 1} \) by \( \epsilon = \gamma / \sqrt{p + 1} \) for all \( |u - \bar{u}| \leq \epsilon \). For simplicity, let \( C_\zeta := 1 + (\zeta_\lambda^2 + \zeta_\lambda^2)/(1 + p) \). Since \( \zeta_\lambda = O(\sqrt{p}) \), \( C_\zeta = O(1) \).

Finally, we have \( \| \Delta_{\beta} \|_F^2 + \| \Delta_{L} \|_F^2 / NT \leq 2C_\zeta \gamma^2 \), the following holds for all \( u \in \mathcal{U} \) and all \( |u - \bar{u}| \leq \epsilon \) with probability one,
\[
\| \Delta_{L} + L_{0}(u) - L_{0}(\bar{u}) \|_* \\
\leq \| \Delta_{L} \|_* + \| L_{0}(u) - L_{0}(\bar{u}) \|_* \\
\leq \frac{2(1 + C_\lambda)}{C_\lambda} \| \mathcal{P}(u) \Delta_{L}(u) \|_* + C_{\text{Come}} \sqrt{p \log((p + 1)NT)(N \wedge T)} \| \Delta_{\beta} \|_F + \sqrt{2r} \zeta_\lambda \sqrt{NT} \gamma / \sqrt{p + 1}
\]

33
\[ \left( \frac{2(1 + C_\lambda) \sqrt{6C_\zeta}}{C_\lambda} + \sqrt{2}\zeta_L \right) \sqrt{NT\bar{r}\gamma} + C_{cone} \sqrt{p \log((p + 1)NT)(N \wedge T)} \sqrt{2C_\zeta\gamma} \]

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}} = \left\{ \Delta_L \in \mathbb{R}^{N \times T} : \right\}

\[ \|\Delta_L\| \leq \left( \frac{2(1 + C_\lambda) \sqrt{6C_\zeta}}{C_\lambda} + \sqrt{2}\zeta_L \right) \sqrt{NT\bar{r}\gamma} + C_{cone} \sqrt{2C_\zeta p \log((p + 1)NT)(N \wedge T)\gamma} \].

Therefore, in the intersection of the ball \( \|\Delta_\beta\|^2_F + \|\Delta_L\|^2_F / NT \leq 2C_\zeta \gamma^2 \) and \( \mathcal{R}_u \) for all \( u \in \mathcal{U} \), the matrices \( \Delta_L \), \( \Delta_L + L_0(u) - L_0(\bar{u}) \) 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,

\[ C^0(\gamma) \leq 2 \cdot \sup_{\Delta_L, u \in \mathcal{U}, \Delta_L \in \bar{\mathcal{R}}} \left| G^0(\delta_{it}(X_{it}^u \Delta_\beta + \Delta_{L,it}, \bar{u})) \right| := 2C^1(\gamma). \quad \text{(S.B.65)} \]

Define the following event:

\[ \Omega_2 := \left\{ \sup_{\Delta_L, u \in \mathcal{U}} \left| G^0(\Delta_{L,it}) \right| \right\}

\[ \leq C_{sp} \left[ \left( C_{cone} \sqrt{2C_\zeta p \log((p + 1)NT)} \right) \vee \left( \sqrt{N \wedge T} \left( \frac{\sqrt{6C_\zeta 2(1 + C_\lambda)}}{C_\lambda} + \sqrt{2}\zeta_L \right) \sqrt{\bar{r}} \right) \right] \gamma \].

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

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

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

\[ \mathbb{E} \left[ e^{\tau' C^1(\gamma)} | W_X, \Omega_1 \right] \]

\[ \leq \frac{\sqrt{p + 1}}{\gamma} \max_{u \in \mathcal{U}} \mathbb{E} \left[ \exp \left( \tau' \sup_{\Delta_L \in \bar{\mathcal{R}}} \left| G^0(\delta_{it}(X_{it}^u \Delta_\beta + \Delta_{L,it}, \bar{u})) \right| W_X, \Omega_1 \right) \right] \]
\begin{align*}
& \leq \frac{\sqrt{p+1}}{\gamma} \mathbb{E} \left[ \exp \left( 2\tau' \sup_{\|\Delta L\|_2^p \leq 2\gamma^2} |G^0(X'_{it}\Delta_{it} + \Delta L_{it})| \right) \right] W_X, \Omega_1 \\
& \leq \frac{\sqrt{p+1}}{\gamma} \mathbb{E} \left[ \exp \left( 2\tau' \sup_{\|\Delta L\|_2^p \leq 2\gamma^2} |G^0(X'_{it}\Delta_{it})| \right) + 2\tau' \sup_{\|\Delta L\|_2^p \leq 2\gamma^2} |G^0(\Delta L_{it})| \right) W_X, \Omega_1 \\
& \leq \frac{\sqrt{2}(p+1)}{\gamma} \mathbb{E} \left[ \exp \left( 4\tau' \sup_{\|\Delta L\|_2^p \leq 2\gamma^2} |G^0(X'_{it}\Delta_{it})| \right) \right] \mathbb{E} \left[ \exp \left( 4\tau' \sup_{\|\Delta L\|_2^p \leq 2\gamma^2} |G^0(\Delta L_{it})| \right) \right] \frac{1}{2}, \quad \text{(S.B.67)}
\end{align*}

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 \( B_0(\gamma) \).

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

\[ \mathbb{E} \left[ \exp \left( 4\tau' \sup_{\|\Delta L\|_2^p \leq 2\gamma^2} |G^0(\Delta L_{it})| \right) \right] \leq \mathbb{P}(\Omega_2) \exp \left[ 4\tau' C_{sp} \left( C_{cone} \sqrt{2C_p \log((p+1)NT)} \vee \left( \sqrt{N \vee T} \left( \frac{\sqrt{6}C_2(1+C_{X})}{C_{X}} + \frac{\sqrt{2}}{\sqrt{1}} \right) \frac{\sqrt{7}}{7} \right) \right) \right] + 2 \exp(-(N \vee T)) \exp \left( 4\tau' \left[ C_{cone} \sqrt{2C_p \log((p+1)NT)(N \vee T)} \left( \frac{\sqrt{6}C_2(1+C_{X})}{C_{X}} + \frac{\sqrt{2}}{\sqrt{1}} \right) \frac{\sqrt{7}}{7} \right) \right).

Let

\[ \tau' = \frac{\sqrt{\log(NT)}}{4\sqrt{2} \left[ \sqrt{2C_p \log((p+1)NT)(C_X \vee C_{cone}^2) + C_{sp} \sqrt{N \vee T} \left( \frac{\sqrt{6}C_2(1+C_{X})}{C_{X}} + \frac{\sqrt{2}}{\sqrt{1}} \right) \frac{\sqrt{7}}{7} \right]} \frac{\sqrt{2}}{2}. \]

Then \( \mathbb{E} \left[ \exp \left( 4\tau' \sup_{\|\Delta L\|_2^p \leq 2\gamma^2} |G^0(\Delta_{it})| \right) \right] \leq 2 \exp \left( \sqrt{\log(NT)} \right) \leq 2 \exp(\log(NT)) = 2\sqrt{NT} \) for large enough \( N \) and \( T \). Substituting it into (S.B.67), we have \( \mathbb{E} \left[ e^{\tau'C_1(\gamma)} \right] \leq 2(p+1)\sqrt{NT}/\gamma \). Let

\[ K_3 = \frac{8\sqrt{2} \left[ \sqrt{2C_p \log((p+1)NT)(C_X \vee C_{cone}^2) + C_{sp} \sqrt{N \vee T} \left( \frac{\sqrt{6}C_2(1+C_{X})}{C_{X}} + \frac{\sqrt{2}}{\sqrt{1}} \right) \frac{\sqrt{7}}{7} \right]} \frac{\sqrt{2}}{2}. \]
In view of (S.B.66), we thus obtain
\[
\mathbb{P} \left( C^0(\gamma) \geq 2K_3 \bigg| \Omega_1 \right) = \mathbb{E} \left[ \mathbb{P} \left( C^0(\gamma) \geq 2K_3 \big| W_X, \Omega_1 \right) \big| \Omega_1 \right] \leq \frac{2(p + 1)}{\gamma (NT)^{3/2}} \to 0, \quad (\text{S.B.68})
\]
where the equality is by the law of iterated expectation. Convergence follows the definition of \( \gamma \) and \( p = o((N \wedge T)/(\log(NT)\alpha^2_{NT})) \), where the latter is implied by Assumption 3.

Now, recall \( s = C_{sup} \sqrt{\log(NT)} \left( \sqrt{p \log((p + 1)NT)} \vee \sqrt{r(N \wedge T)} \right) \gamma \) and let
\[
C_{sup} = 192\sqrt{2} \left( \sqrt{2(C_X \vee C_{Sp}^2 C_{cone})C_{s}} + C_{Sp} \left( \frac{\sqrt{6C_3^2(1 + C_\lambda)} + \sqrt{2}\zeta_{L}}{C_\lambda} \right) \right)
\]
\[
= 192\sqrt{2} \left( 2 \sqrt{1 \vee \left( \frac{50C_3^2 C_{sp}^2}{C_\lambda^2 C_{op}^2} \right)} C_{\zeta} C_X + C_{Sp} \left( \frac{\sqrt{6C_3^2(1 + C_\lambda)} + \sqrt{2}\zeta_{L}}{C_\lambda} \right) \right),
\]
where the second equality is by \( C_{cone} = 5\sqrt{2C_X/C_\lambda C_{op}} \). We can see that \( C_{sup} \) decreases as \( C_\lambda \) 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
\[
\mathbb{P} \left( A^0(\gamma) \geq \frac{s}{4} \bigg| \Omega_1 \right) \leq 3 \max \left\{ \mathbb{P} \left( B_1^0(\gamma) \geq \frac{s}{12} \bigg| \Omega_1 \right), \mathbb{P} \left( B_2^0(\gamma) \geq \frac{s}{12} \bigg| \Omega_1 \right), \mathbb{P} \left( C^0(\gamma) \geq \frac{s}{12} \bigg| \Omega_1 \right) \right\}
\]
\[
\leq 3 \max \left\{ \mathbb{P} \left( B_1^0(\gamma) \geq K_1 \bigg| \Omega_1 \right), \mathbb{P} \left( B_2^0(\gamma) \geq K_2 \bigg| \Omega_1 \right), \mathbb{P} \left( C^0(\gamma) \geq 2K_3 \bigg| \Omega_1 \right) \right\} \to 0. \quad (\text{S.B.69})
\]
By equations (S.B.59), (S.B.60) and (S.B.69), we obtain \( \mathbb{P}(A(\gamma) \geq s) \to 0. \)

**Proof of Lemma S.B.6**

We have
\[
\int_0^{w_2} \left( \mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0) \right) dz - \int_0^{kw_2} \left( \mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0) \right) dz = \int_{kw_2}^{w_2} \left( \mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0) \right) dz.
\]
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 \( z \in [kw_2, w_2] \) as \( \kappa \in (0, 1] \); the integral on the right side is nonnegative. If \( w_2 < 0, \) \( \int_{kw_2}^{w_2} \left( \mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0) \right) dz = \int_{kw_2}^{w_2} \left( \mathbb{1}(w_1 \leq 0) - \mathbb{1}(w_1 \leq z) \right) dz \) which is again nonnegative as \( \mathbb{1}(w_1 \leq 0) = \mathbb{1}(w_1 \leq z) \) for all \( z \in [w_2, kw_2] \).
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 $\mathcal{D}^{(2)} = \mathbb{R}^p \times \{ \Delta_L \in \mathbb{R}^{N \times T} : \| P_\Omega \Delta_L \|_F^2 \geq C_{sm} \| \Delta_L \|_F^2 \}$. Let $\gamma = C_{error,2} \alpha_{NT}^2 \left( (1 + C_\lambda) \vee \log(NT) \right) \left( \sqrt{p \log((p + 1)NT)/NT} \vee \sqrt{r/(N \wedge T)} \right)$ 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

$$
\mathbb{E} \left[ \rho_u \left( V(u) - \sum_{j=1}^p X_j \Delta_{\beta,j} - \Delta_L \right) - \rho_u(V(u)) \right] \\
= \sum_{i,t} \mathbb{E} \left[ \int_0^{\Delta_{L,it} + X_{it}' \Delta_\beta} (F_{it(u)}|W_X(s) - F_{it(u)}|W_X(0)) \, ds \right] \\
\geq \sum_{i,t} \mathbb{E} \left[ \mathbb{1} \left( X_{it}' X_{it} \leq \alpha_{NT}/\gamma \right) \int_0^{\Delta_{L,it} + X_{it}' \Delta_\beta} (F_{it(u)}|W_X(s) - F_{it(u)}|W_X(0)) \, ds \right] \\
\nonumber \\
\geq \sum_{i,t} \sum_{|\Delta_{L,it}| \leq 2 \alpha_{NT}} \mathbb{E} \left[ \mathbb{1} \left( X_{it}' X_{it} \leq \alpha_{NT}/\gamma \right) \int_0^{\Delta_{L,it} + X_{it}' \Delta_\beta} (F_{it(u)}|W_X(s) - F_{it(u)}|W_X(0)) \, ds \right] \\
+ \sum_{i,t} \sum_{|\Delta_{L,it}| > 2 \alpha_{NT}} \mathbb{E} \left[ \mathbb{1} \left( X_{it}' X_{it} \leq \alpha_{NT}/\gamma \right) \int_0^{\Delta_{L,it} + X_{it}' \Delta_\beta} (F_{it(u)}|W_X(s) - F_{it(u)}|W_X(0)) \, ds \right] \\
\geq \frac{(1 + \delta)^2}{18 \alpha_{NT}^2} \sum_{i,t} \mathbb{E} \left[ \mathbb{1} \left( X_{it}' X_{it} \leq \alpha_{NT}/\gamma \right) \right] \frac{X_{it}' \Delta_\beta + (P_\Omega \Delta_L)_{it}}{2} \\
\geq \frac{(1 + \delta)^2}{18 \alpha_{NT}^2} \left( 0.5 \sum_{i,t} \mathbb{E} \left[ \mathbb{1} \left( X_{it}' X_{it} \leq \alpha_{NT}/\gamma \right) \right] (X_{it}' \Delta_\beta + (P_\Omega \Delta_L)_{it})^2 \right) - \| P_\Omega \Delta_L \|_F^2 \\
\geq \frac{(1 + \delta)^2}{18 \alpha_{NT}^2} \left( 0.5 \sum_{i,t} \mathbb{E} \left[ \mathbb{1} \left( X_{it}' X_{it} \leq \alpha_{NT}/\gamma \right) \right] (X_{it}' \Delta_\beta + (P_\Omega \Delta_L)_{it})^2 \right) - (1 - C_{sm}) \| \Delta_L \|_F^2 \\
\geq \frac{0.5 C_{min}}{\alpha_{NT}^2} \left( \mathbb{E} \left[ \left( \sum_{j=1}^p X_j \Delta_{\beta,j} + \Delta_L \right)^2 \right] - (\epsilon_{NT}(NT \| \Delta_\beta \|_F^2 + \| \Delta_L \|_F^2) + 2(1 - C_{sm}) \| \Delta_L \|_F^2) \right)
$$

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 $\mathbb{1} \left( X_{it}' X_{it} \leq \alpha_{NT}/\gamma \right)$ is true and when $\Delta_{L,it} > 2 \alpha_{NT}$, we have

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

37
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 $f$ and $\delta$ are as in Assumption 2. The fourth and the fifth inequalities are by $\Delta_{L,lt} = (\mathcal{P}_\Omega \Delta_L)_{lt} + (\mathcal{P}_\Omega + \Delta_L)_{lt}$ 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 $(\beta^{(k+1)}, V^{(k+1)})$

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

where $\mathcal{L}$ 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 $\mathcal{L}(L^{(k+1)}, \beta^{(k+1)}, V^{(k+1)}, H^{(k)})$.

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

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

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

$$\mathcal{L} \left( L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)} \right) \leq \mathcal{L} \left( L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)} \right) - \langle V^{(l)} - V^{(l+1)}, \nabla_V \mathcal{L}(L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)}) \rangle$$

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

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

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

$$\mathcal{L} \left( L^{(k+1)}, \beta^{(l+1)}, V^{(l+1)}, H^{(k)} \right) \leq \mathcal{L} \left( L^{(k+1)}, \beta^{(l)}, V^{(l+1)}, H^{(k)} \right), \forall l.$$
Combining the above two inequalities, we have the following for any initial value \((\beta^{(0)}, V^{(0)})\),
\[
\mathcal{L}(L^{(k+1)}, \beta^{(0)}, V^{(0)}, H^{(k)}) \geq \mathcal{L}(L^{(k+1)}, \beta^{(0)}, V^{(1)}, H^{(k)}) \geq \mathcal{L}(L^{(k+1)}, \beta^{(1)}, V^{(1)}, H^{(k)}) \ldots
\]
Therefore, sequence \((\mathcal{L}(L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)}))\) is decreasing. By the monotone convergence theorem, \(\lim_{l \to \infty} \mathcal{L}(L^{(k+1)}, \beta^{(l)}, V^{(l)}, H^{(k)}) \to \mathcal{L}(L^{(k+1)}, \beta^{(k+1)}, V^{(k+1)}, H^{(k)})\) where the limit is the global minimum of \(\mathcal{L}(L^{(k+1)}, \cdot, \cdot, H^{(k)})\).

Lemma S.B.7 shows that \((\beta^{(l)}, V^{(l)})\) converges to \((\beta^{(k+1)}, V^{(k+1)})\), a global minimizer of \(\mathcal{L}(L^{(k+1)}, \cdot, \cdot, H^{(k)})\). 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 \(\mu\). Here our \(\mu\) is fixed to simplify computation. Having a fixed \(\mu\) requires a new argument especially in the proof of Theorem S.B.1. Define
\[
\tilde{H}^{(k+1)} \equiv H^{(k)} - \mu \left( V^{(k)} + \sum_{j=1}^{p} X_j \beta_j^{(k)} + L^{(k+1)} - Y \right) \quad \text{(S.B.70)}
\]
Combining (A.5) and (S.B.70) and defining \(\hat{V} := Y - \hat{L} - \sum_{j=1}^{p} X_j \hat{\beta}_j\) where \((\hat{\beta}, \hat{L})\) are our estimator \((u\) is suppressed for the ease of notation\), we have
\[
H^{(k+1)} - H^{(k)} = \mu \left( Y - V^{(k+1)} - \sum_{j=1}^{p} X_j \beta_j^{(k+1)} - L^{(k+1)} \right)
\]
\[
= \mu \left( \hat{V} - V^{(k+1)} \right) + \sum_{j=1}^{p} X_j \left( \hat{\beta}_j - \beta_j^{(k+1)} \right) + \left( \hat{L} - L^{(k+1)} \right) \quad \text{(S.B.71)}
\]
and
\[
\tilde{H}^{(k+1)} - H^{(k+1)} = \mu \left( (V^{(k+1)} - V^{(k)}) + \sum_{j=1}^{p} X_j \left( \beta_j^{(k+1)} - \beta_j^{(k)} \right) \right) \quad \text{(S.B.72)}
\]
Meanwhile, consider the Lagrangian without the penalty:
\[
\hat{\mathcal{L}}(L, \beta, V, H) = \frac{1}{\lambda NT} \rho_u(V) + \|L\|_s + \left( H, Y - \sum_{j=1}^{p} X_j \beta_j - L - V \right) \quad \text{(S.B.73)}
\]
Note that this is the Lagrangian of the minimization problem that defines our estimator \((\hat{\beta}, \hat{L})\). Let \(\tilde{H}\) 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 $j = 1, \ldots, p$ and for all $k \geq 1$, we have

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

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

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

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

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

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

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

$$+ \mu^{-2} \left[ \left\| H^{(k)} - \hat{H} \right\|_F^2 - \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 \right]$$

$$- 2\mu^{-1} \left( \left\langle V^{(k+1)} - V^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle + \left\langle L^{(k+1)} - \hat{L}, H^{(k+1)} - \hat{H} \right\rangle + \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle \right).$$

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

$$\mu^{-1} \left\langle H^{(k+1)} - H^{(k)}, H^{(k+1)} - \hat{H} \right\rangle$$

$$= - \left\langle V^{(k+1)} - \hat{V} + \sum_{j=1}^{p} X_j (\beta_j^{(k+1)} - \hat{\beta}_j), H^{(k+1)} - \hat{H} \right\rangle - \left\langle L^{(k+1)} - \hat{L}, H^{(k+1)} - \hat{H} \right\rangle$$

$$= - \left\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \right\rangle - \left\langle L^{(k+1)} - \hat{L}, H^{(k+1)} - \hat{H} \right\rangle$$

$$+ \mu \left\langle L^{(k+1)} - \hat{L}, (V^{(k+1)} - V^{(k)}) + \sum_{j=1}^{p} X_j (\beta_j^{(k+1)} - \beta_j^{(k)}) \right\rangle. \hfill (S.B.76)$$

Therefore,

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

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

40
Proof. By the optimality of $g$ is a decreasing sequence in the equality above are all nonnegative. That proves the left hand side of the first equality above is due to the fact that 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 $\nabla f$ is a subgradient of $f$.

Lemma S.B.11.

\[
\lim_{k \to \infty} \|H^{(k+1)} - H^{(k)}\|_F^2 = 0,
\]
\[
\lim_{k \to \infty} \langle V^{(k+1)} - \hat{V}, H^{(k-1)} - \hat{H} \rangle = 0,
\]
\[
\lim_{k \to \infty} \langle L^{(k+1)} - \hat{L}, H^{(k+1)} - \hat{H} \rangle = 0.
\]

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

\[
\hat{H} \in \nabla \|\hat{L}\|_*, \quad \hat{H} \in \frac{1}{\lambda NT} \nabla \rho_u(\hat{V}). \quad (\text{S.B.77})
\]
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
\[
\tilde{H}^{(k+1)} \in \nabla \| L^{(k+1)} \|_*, \quad H^{(k+1)} \in \frac{1}{\lambda NT} \nabla \mu_u(V^{(k+1)}).
\] (S.B.78)

Therefore, by Lemma S.B.10 we have the following inequalities:
\[
\begin{align*}
&\langle V^{(k+1)} - V^{(k)}, H^{(k+1)} - H^{(k)} \rangle \geq 0, \\
&\langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \rangle \geq 0, \\
&\langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \rangle \geq 0.
\end{align*}
\]

Now by the nonnegativeness of the three inner products and by Lemma S.B.9, we have that \( \Psi^{(k)} := \| (V^{(k+1)} - \hat{V}) + \sum_{j=1}^p X_j (\beta_j^{(k+1)} - \hat{\beta}_j) \|^2_F + \mu^{-2} \| H^{(k+1)} - \hat{H} \|^2_F \) 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
\[
0 \leq \mu^{-2} \| H^{(k+1)} - H^{(k)} \|^2_F + \mu^{-1} \langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \rangle + \mu^{-1} \langle L^{(k+1)} - \hat{L}, \tilde{H}^{(k+1)} - \hat{H} \rangle \leq \Psi^{(k)} - \Psi^{(k+1)} \to 0.
\]

Hence, the desired results obtain since \( \| H^{(k+1)} - H^{(k)} \|^2_F, \langle V^{(k+1)} - \hat{V}, H^{(k+1)} - \hat{H} \rangle \) 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)*:

\[
\begin{align*}
\lim_{k \to \infty} \left\| Y - L^{(k)} - \sum_{j=1}^p X_j \beta_j^{(k)} - V^{(k)} \right\|^2_F &= 0, \\
\lim_{k \to \infty} \frac{1}{\lambda NT} \mu_u \left( Y - \sum_{j=1}^p X_j \beta_j^{(k)} - L^{(k)} \right) + \left\| L^{(k)} \right\|_* &= \frac{1}{\lambda NT} \mu_u \left( Y - \hat{L} - \sum_{j=1}^p X_j \beta_j \right) + \left\| \hat{L} \right\|_*.
\end{align*}
\] (S.B.79)
(S.B.80)

**Proof.** Since \( \| H^{(k)} - H^{(k-1)} \|^2_F \to 0 \) as \( k \to \infty \) by Lemma S.B.11, under equation (S.B.71),
\[
\left\| Y - L^{(k)} - \sum_{j=1}^p X_j \beta_j^{(k)} - V^{(k)} \right\|^2_F = \mu^{-1} \left\| H^{(k)} - H^{(k-1)} \right\|^2_F \to 0.
\] (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 \( \tilde{H}^{(k)} \in \nabla \|L^{(k)}\|_\star \) and that \( H^{(k)} \in \frac{1}{\lambda NT} \nabla \rho_u(V^{(k)}) \) by equation (S.B.78). By the definition of subgradient,

\[
\frac{1}{\lambda NT} \rho_u \left( V^{(k)} \right) + \left\| L^{(k)} \right\|_\star \\
\leq \frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_\star - \left\langle \hat{L} - L^{(k)}, \tilde{H}^{(k)} \right\rangle - \left\langle \hat{V} - V^{(k)}, H^{(k)} \right\rangle \\
= \frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_\star + \left\langle \hat{L} - L^{(k)}, \hat{H} - \tilde{H}^{(k)} \right\rangle + \left\langle \hat{V} - V^{(k)}, \hat{H} - H^{(k)} \right\rangle \\
- \left\langle \hat{L} + \hat{V} - L^{(k)} - V^{(k)}, \hat{H} \right\rangle \\
= \frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_\star + \left\langle \hat{L} - L^{(k)}, \hat{H} - \tilde{H}^{(k)} \right\rangle + \left\langle \hat{V} - V^{(k)}, \hat{H} - H^{(k)} \right\rangle \\
- \left\langle Y - \sum_{j=1}^p X_j \hat{\beta}_j - L^{(k)} - V^{(k)}, \hat{H} \right\rangle \\
\leq \frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_\star + \left\langle \hat{L} - L^{(k)}, \hat{H} - \tilde{H}^{(k)} \right\rangle + \left\langle \hat{V} - V^{(k)}, \hat{H} - H^{(k)} \right\rangle \\
+ \left\| \hat{H} \right\|_F \cdot \left\| Y - \sum_{j=1}^p X_j \beta_j - L^{(k)} - V^{(k)} \right\|_F \\
=: \frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_\star + a_k. \tag{S.B.82}
\]

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

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

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
\[
\left| \frac{1}{\lambda NT} \rho_u \left( V^{(k)} \right) + \left\| L^{(k)} \right\|_* - \frac{1}{\lambda NT} \rho_u \left( \tilde{V}^{(k)} \right) - \left\| L^{(k)} \right\|_* \right| \leq b_k \rightarrow 0 \quad \text{(S.B.83)}
\]

Therefore,
\[
\frac{1}{\lambda NT} \rho_u \left( V^{(k)} \right) + \left\| L^{(k)} \right\|_* \geq \frac{1}{\lambda NT} \rho_u \left( \tilde{V}^{(k)} \right) + \left\| L^{(k)} \right\|_* - b_k \geq \frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_* - b_k \quad \text{(S.B.84)}
\]

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 \(\beta\) satisfying the constraint \(L + V + \sum_{j=1}^{p} X_j \beta_j = Y, \rho_u(V)/(\lambda NT) + \|L\|_*\) is minimized at \((\hat{V}, \hat{L})\). Combining equations (S.B.82) and (S.B.84), we have

\[
\frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_* - b_k \leq \frac{1}{\lambda NT} \rho_u \left( V^{(k)} \right) + \left\| L^{(k)} \right\|_* \leq \frac{1}{\lambda NT} \rho_u \left( \hat{V} \right) + \left\| \hat{L} \right\|_* + a_k.
\]

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

\[
\lim_{k \rightarrow \infty} \frac{1}{\lambda NT} \rho_u \left( V^{(k)} \right) + \left\| L^{(k)} \right\|_* = \frac{1}{\lambda NT} \rho_u \left( Y - \sum_{j=1}^{p} X_j \hat{\beta} - \hat{L} \right) + \left\| \hat{L} \right\|_* . \quad \text{(S.B.85)}
\]

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**


