Regularized Quantile Regression with Interactive Fixed Effects *

Junlong Feng†

This version: November 27, 2019

Abstract

I consider nuclear norm penalized quantile regression for large N and large T panel data models with interactive fixed effects. The estimator solves a convex minimization problem, not requiring pre-estimation of the (number of the) fixed effects. Uniform rates are obtained for both the regression coefficients and the common component estimators. The rate of the latter is nearly optimal. To derive the rates, I also show new results that establish uniform bounds related to random matrices of jump processes. These results may have independent interest. Finally, I conduct Monte Carlo simulations to illustrate the estimator's finite sample performance.

Keywords: Quantile regression, interactive fixed effects, nuclear norm, regularized regression, high-dimensional data.

1 Introduction

Panel data models are widely applied in economics and finance. Allowing for rich heterogeneity, interactive fixed effects are important components in such models in a lot of applications. Since fixed effects may not only impact the conditional mean of the outcome variable, but also have heterogeneous effects on its distribution, quantile regression would be

*I thank Jushan Bai, Roger Koenker, Simon Lee, Roger Moon, Serena Ng, José Luis Montiel Olea, Bernard Salanié, Jörg Stoye and participants at the Columbia Econometrics Colloquium for helpful comments and discussions. All errors are my own.

†Department of Economics, Columbia University; junlong.feng@columbia.edu.

handy in such cases. However, in contrast to the well-studied mean regression (e.g. Pesaran (2006), Bai (2009) and Moon and Weidner (2015)), quantile regression for panel data with interactive fixed effects only received attention recently (e.g. Harding and Lamarche (2014), Ando and Bai (2019), Chen, Dolado and Gonzalo (2019) and Chen (2019)).

In this paper, I propose nuclear norm regularized quantile regression for panel data models with interactive fixed effects. The estimator solves a convex problem, and I allow the fixed effect structure to vary across quantiles. Meanwhile, I do not need to pre-estimate the (number of the) fixed effects. To highlight the departure from the existing approaches, let us think about a simple location shift model with exogenous regressors and i.i.d. data as an illustration. Let $Y_{it} = X_{it}'\beta_0^o + F_t^{o\prime}\Lambda_i^o + \varepsilon_{it} \equiv X_{it}'\beta_0(u) + F_t^{o\prime}\Lambda_i^o + [q_\varepsilon(U_{it}) - q_\varepsilon(u)]$, where F_t^o and Λ_i^o contain r fixed effects ($r < \min\{N, T\}$), $q_\varepsilon(\cdot)$ is the quantile function of ε_{it}, and $U_{it} \sim \text{Unif}[0, 1]$. In the papers mentioned above, though they all have different focuses, they essentially estimate $(\beta_0(u), F_t^o, \Lambda_i^o)$ by minimizing the check function: $\sum_{i,t}\rho_u(Y_{it} - X_{it}\beta - F_t'\Lambda_i)$. To implement, r needs to be pre-estimated. As this objective function is nonconvex in (β, F_t, Λ_i), we may end up with a local minimum.

To resolve the issue of nonconvexity, note that the check function is indeed convex in $(\beta, F_t'\Lambda_i)$. Hence, by treating the interactive fixed effects as single parameters instead, we obtain a convex problem: $\min_{(\beta,\{L_{it}\})}\sum_{it}\rho_u(Y_{it} - X_{it}'\beta - L_{it})$.The minimizer of this problem is trivially $(0, \{Y_{it}\})$ if we do not regulate L. Denote $L_0^o = F_t^{o\prime}\Lambda_i^o$. By construction, $\text{rank}(L_0^o) \leq r < \min\{N, T\}$. The low-rankness of the true parameter thus motivates a penalty term added to the check function that regularizes the rank of L.

The penalty function of choice is the nuclear norm of a matrix $||\cdot||_*$, i.e., the sum of a matrix's singular values. As all of them are nonnegative by definition, the nuclear norm is equivalent to the ℓ_1 norm of the vector of the singular values. Recall that the rank of a matrix is the number of nonzero singular values. Therefore, the nuclear norm penalty is analogous to the LASSO penalty in high dimensional regression; as the latter is the tightest convex relaxation of the number of the nonzero coefficients, the former is the counterpart for the rank of a matrix.

The practical benefits of the nuclear norm regularized quantile regression are two-fold. Unlike the existing methods mentioned earlier, it is unnecessary to know r so no pre-estimation is needed. The rank of the estimated common component is partly determined by the weight assigned to the penalty. Also, since the nuclear norm is convex too, the regularized minimization problem is still convex in (β, L). Hence, there is no concern of local minima.

In this paper, I provide uniform rate of convergence for the estimator $(\hat{\beta}(u), \hat{L}(u))$ in $u \in \mathcal{U}$, a compact interval in the interior of $[0, 1]$. The rate of $\hat{L}(u)$ is nearly optimal but the rate of $\hat{\beta}(u)$ is slower than optimal due to the bias caused by regularization. To obtain the rate, the subgradient matrix of the check function of the error processes plays a key role. I prove new results that establish a uniform upper bound for the operator norm of this matrix, and a uniform Hoeffding-type bound for its inner product with other matrices. Also, I develop a new theoretical argument such that the conditional density of Y_{it} only needs to be away from 0 at the true parameter. This is standard in the classical quantile regression literature but weaker than the conditions in Ando and Bai (2019) and Chen, Dolado and Gonzalo (2019), where the density is assumed to be bounded away from 0 in compact subsets of the support. These results may have independent interest. I also discuss post-regularization procedures where a consistent estimator for r is provided as a by-product.

This paper adds to the literature of quantile regression for panel data. Since Koenker (2004), panel data quantile regression began to draw increasing attention. Abrevaya and Dahl (2008), Lamarche (2010), Canay (2011), Kato, Galvao Jr and Montes-Rojas (2012), Galvao, Lamarche and Lima (2013) and Galvao and Kato (2016) study quantile regression with one-way or two-way fixed effects. Harding and Lamarche (2014) considers interactive fixed effects with endogenous regressors. They require F_t^o to be pre-estimated or known. Chen, Dolado and Gonzalo (2019) prospses a quantile factor model without the regressors. They estimate the factors and the factor loadings via nonconvex minimizations. Pre-estimation of the number of the factors is needed. Ando and Bai (2019) considers quantile regression with heterogeneous coefficients. They propose both a frequentist and a Bayesian estimation procedure. The number of the factors also needs to be estimated first, and the minimization problem is noncovex. Chen (2019) proposes a two-step estimator: by assuming the common factors are not quantile-dependent, they are estimated by the principal component analysis in the first step, and the regression coefficients and the individual fixed effects are estimated in the second step via smoothed quantile regression. Chao, Härdle and Yuan (2019) considers nuclear norm penalized multi-task quantile regression where multiple outcome variables are simultaneously considered and the coefficient matrix is low rank.

Another literature this paper speaks to is the nuclear norm regularized estimation. This literature was initially motivated by low rank matrix completion or recovery in compressed sensing and other applications in computer science, etc, for instance Candès and Recht (2009), Ganesh et al. (2010), Zhou et al. (2010), Candès et al. (2011), Hsu, Kakade and Zhang (2011), Negahban and Wainwright (2011), Agarwal, Negahban and Wainwright (2012) and Negahban et al. (2012) among others. The main parameter of interest in this literature is the low-rank matrix and they require the error terms to be either nonstochastic or to have finite second moments. Bai and Feng (2019) allows the stochastic error to be non-sparse and fat-tailed; the existence of its moments are not required. Nuclear norm regularized estimation and matrix completion related topics have also gained interest in econometrics recently. Bai and Ng (2019*a*) considers using factor analysis to impute missing data and counterfactuals. Bai and Ng (2019*b*) considers regularized estimation for approximate factor models with singular values thresholding. Athey et al. (2018), Moon and Weidner (2019) and Chernozhukov et al. (2019) consider mean regression with interactive fixed effects. In Moon and Weidner (2019), they also briefly discuss nuclear norm regularized quantile regression with a single regressor as an extension. The rate of convergence they obtain is pointwise, and is almost the square root of the rate obtained here. Also their approach focuses on the regression coefficients, but I obtain rates for both the coefficients and the common component.

The rest of the paper is organized as follows. Section 2 introduces the model and the estimator. Section 3 discusses the *restricted set*, an important theoretical device that is useful to establish the main results. The main results are presented in Section 4. Section 5 shows a Monte Carlo simulation experiment. Section 6 concludes. The proofs of the results in Sections 3 and 4 and the technical lemmas are collected in Appendices A, B and C respectively.

Notation

Besides the nuclear norm $||\cdot||_*$, I will use four additional matrix norms. Let $||\cdot||$, $||\cdot||_F$, $||\cdot||_1$, and $||\cdot||_\infty$ denote the operator norm, the Frobenius norm, the ℓ_1 norm and the maximum norm. For two generic scalars, $a \vee b$ and $a \wedge b$ return the max and the min between a and b, respectively. For a generic random sample $W_1, ..., W_{NT}$ and a function f, denote the empirical process by $\mathbb{G}_{NT}(f) \equiv \mathbb{G}_{NT}(f(W_{it})) = \frac{1}{\sqrt{NT}} \sum_{i,t}(f(W_i) - \mathbb{E}(f(W_i)))$ where $\mathbb{E}$ denotes the expectation operator. Finally, I use the notion *"with high probability"* when an event occurs with probability arbitrarily close to 1 for large enough N and T.

2 The Model and the Estimator

I consider the following model for a panel dataset $(Y_{it}, X_{it} : i \in \{1, ..., N\}, t \in \{1, ..., T\})$:

$$Y_{it} = X_{it}'\beta_0(U_{it}) + \sum_{k=1}^{\bar{r}} \mathbb{1}_k(U_{it}) F_{kt}(U_{it}) \Lambda_{ki}(U_{it}), \tag{2.1}$$

where X_{it} is a $p \times 1$ vector of covariates, $U_{it} \sim \text{Unif}[0, 1]$ is unobservable, $F_t(U_{it})$ and $\Lambda_i(U_{it})$ are latent fixed effects with the total number $\bar{r}$ unknown. The interactive fixed effects part is similar to that in Ando and Bai (2019) and Chen, Dolado and Gonzalo (2019). Both the fixed effects and the effective number of them are allowed to vary in U_{it}.

The key observation of model (2.1) is that the interactive fixed effects form a low-rank $N \times T$ matrix at a fixed u if $\bar{r}$ is small relative to N and T. To see this, let us rewrite equation (2.1) as follows:

$$
\begin{aligned}
Y_{it} =& X_{it}'\beta_0(u) + \underbrace{\sum_{k=1}^{\bar{r}} \mathbb{1}_k(u) F_{kt}(u)\Lambda_{ki}(u)}_{\equiv L_{0,it}(u)} \\
&+ \underbrace{X_{it}'\Big(\beta_0(U_{it}) - \beta_0(u)\Big) + \sum_{k=1}^{\bar{r}} \Big(\mathbb{1}_k(U_{it}) F_{kt}(U_{it})\Lambda_{ki}(U_{it}) - \mathbb{1}_k(u) F_{kt}(u)\Lambda_{ki}(u)\Big)}_{\equiv V_{it}(u)}
\end{aligned}
$$

Or more compactly,

$$
Y = \sum_{j=1}^{p} X_j \beta_{0,j}(u) + L_0(u) + V(u) \tag{2.2}
$$

where $\text{rank}(L_0(u)) \le r(u) \le \bar{r}$ and X_j $(N \times T)$ is the jth regressor. The number of regressors p and the number of factors $\bar{r}$ are fixed for simplicity but are potentially allowed to grow with N and T. Denote the conditional quantile of V_{it} at u by $q_{V_{it}|X_{it},F_t}(u)$. I assume the function $X_{it}'\beta(\cdot) + \sum_{k=1}^{\bar{r}} \mathbb{1}_k(\cdot) F_{kt}(\cdot)\Lambda_{kt}(\cdot)$ is strictly increasing, then we have $q_{V_{it}|X_{it},F_t}(u) = 0$ by construction. The following are two models that admit the representations (2.1) and (2.2).

Example 1 (Location Shift Model)**.** *This is the model we've seen in the introduction.* $Y_{it} = X_{it}'\beta^o + F_t^{o'}\Lambda_i^o + \epsilon_{it}$. *Let* $q_\epsilon(\cdot)$ *be the quantile function of* ϵ, *then* $Y_{it} = X_{it}'\beta^o + F_t^{o'}\Lambda_i^o + q_\epsilon(U_{it})$. *At* u, $q_\epsilon(u)$ *is absorbed in the constant.*

Example 2 (Location-Scale Model)**.** $Y_{it} = X_{it}'\beta_0^a + F_t^{o'}\Lambda_t^a + (X_{it}'\beta^b + F_t^{o'}\Lambda_i^b)\epsilon_{it}$. *Let* $q_\epsilon(\cdot)$ *be the quantile function of* ϵ, *then we can rewrite the model as*

$$
Y_{it} = X_{it}'[\beta_0^a + \beta_0^b q_\epsilon(u)] + F_t^{o'}[\Lambda_i^a + \Lambda_i^b q_\epsilon(u)]
$$

where $x'\beta^b + \phi'\Lambda_i^b > 0$ *for all* i *and all* (x, ϕ) *in the support set of* X_{it} *and* F_t. *In this model, both the coefficients on the regressors and the individual fixed effects are functions of* u. *If for some* u, *there are elements in* $\Lambda_i^a + \Lambda_i^b q_\varepsilon(u)$ *equal to* 0, *then the number of the effective fixed effects is smaller than that at other* u. *Hence,* $r(\cdot)$ *depends on* u.

I estimate $(\beta_0(u), L_0(u))$ by the following regularized quantile regression:

$$
(\hat{\beta}(u), \hat{L}(u)) \equiv \arg \min_{\beta \in \mathbb{R}^p, ||L||_\infty \le \alpha} \quad \frac{1}{NT}\boldsymbol{\rho}_u(Y - \sum_{j=1}^{p} X_j\beta_j - L) + \lambda ||L||_* \tag{2.3}
$$

where for a generic matrix Z, $\boldsymbol{\rho}_u(Z) \equiv \sum_{i,t} \rho_u(Z_{it}) \equiv \sum_{i,t} Z_{it}(u - \mathbb{1}_{Z_{it} \leq 0})$ is the sum of the check functions applied to each element in Z. $\lambda \to 0$ and $\alpha \to \infty$ are two tuning parameters that will be specified later.

3 The Restricted Set

To establish consistency, intuitively we hope that $\hat{L}(u)$ lies near the space containing $L_0(u)$. In particular, the nuclear norm penalty is effective if projecting $\hat{L}(u)$ onto the space of $L_0(u)$ obtains a residual matrix that has small nuclear norm. To formalize the idea, let $L_0(u) = R(u)\Sigma(u)S(u)'$ be a singular value decomposition for $L_0(u)$. Let $\Phi(u)$ be the space of matrices defined by $\Phi(u) \equiv \{M \in \mathbb{R}^{N\times T} : \exists A \in \mathbb{R}^{r(u)\times T} \textit{ and } B \in \mathbb{R}^{N\times r(u)} \textit{ s.t. } M = R(u)A + BS(u)'\}$. The linear projection of a generic $N \times T$ matrix W onto this space is

$$P_{\Phi(u)}W = R(u)R(u)'W + WS(u)S(u)' - R(u)R(u)'WS(u)S(u)'$$

Its orthogonal projection is $P_{\Phi^{\perp}(u)}W = \Big(I_{N\times N} - R(u)R(u)'\Big)W\Big(I_{T\times T} - S(u)S(u)'\Big)$. In principle, we hope $P_{\Phi(u)}\hat{L}(u)$ is sufficiently large in nuclear norm compared to $\hat{L}(u)$.

To formalize the idea, let $\hat{\Delta}_\beta(u) = \hat{\beta}(u) - \beta_0(u)$ and $\hat{\Delta}_L(u) = \hat{L}(u) - L_0(u)$. For some $C_1, C_2 > 0$, define the restricted set as follows:

$$\begin{aligned}\mathcal{R}_u \equiv \Big\{(\Delta_\beta, \Delta_L) :& \Big(\lambda - \frac{C_2\sqrt{N \vee T}}{NT}\Big)||P_{\Phi^{\perp}(u)}\Delta_L||_* \\ &\leq C_1\sqrt{\frac{p\log(NT)}{NT}}||\Delta_\beta||_F + \Big(\lambda + \frac{C_2\sqrt{N\vee T}}{NT}\Big)||P_{\Phi(u)}\Delta_L||_*\Big\}\end{aligned} \tag{3.1}$$

For large enough λ, $(\hat{\Delta}_\beta(u), \hat{\Delta}_L(u)) \in \mathcal{R}_u$ indeed implies that the estimation error $\hat{\Delta}_L(u)$ projected to the orthogonal space $\Phi^{\perp}(u)$ of $L_0(u)$ is at most of the same order of that projected on to $\Phi(u)$. Similar notions of the restricted set can be also seen in such as Negahban and Wainwright (2011) for low rank matrix recovery, Belloni and Chernozhukov (2011) for high-dimensional quantile regression, and Chernozhukov et al. (2019) and Moon and Weidner (2019) for mean regression with interactive fixed effects.

I now show that under the following assumption, $(\hat{\Delta}_\beta(u), \hat{\Delta}_L(u)) \in \mathcal{R}_u$.

Assumption 1. *i)* $(X_{it}, U_{it} : i \in \{1, ..., N\}, t \in \{1, ..., T\})$ *are i.i.d. conditional on* F_t. *ii)* $(X_{it}, F_t) \perp\!\!\!\perp U_{it}$. *iii) All the regressors have finite variances and* $\mathbb{E}(X_{it}X_{it}')$ *is invertible.*

In the paper I only consider the conditional i.i.d. case for simplicity. Potential serial correlations are absorbed into F_t. I also assume the regressors are exogenous. Note the assumption implies that $\max_{1\leq j\leq p}||X_j||_F^2 \leq C\cdot NT$ for some constant $C>0$ with high probability.

Lemma 1. *Under Assumption 1, $(\hat{\Delta}_\beta(u), \hat{\Delta}_L(u)) \in \mathcal{R}_u$ uniformly in $u \in \mathcal{U}$ with high probability.*

Proof. See Appendix A. □

It is worth noting that a similar inequality as in the definition of $\mathcal{R}_u$ (3.1) also holds by replacing the nuclear norm with the Frobenius norm, following from the low-rankness of $L_0(u)$. To see this, by the definition of $P_{\Phi(u)}$, $\text{rank}(P_{\Phi(u)}A) \leq 3r(u) \leq 3\bar{r}$ for a generic $N\times T$ matrix A. Since $||A||_F \leq ||A||_* \leq \sqrt{\text{rank}(A)}||A||_F$, the following holds for any $(\Delta_\beta, \Delta_L)\in\mathcal{R}_u$:

$$\Big(\lambda - \frac{C_2\sqrt{N\vee T}}{NT}\Big)||P_{\Phi^\perp(u)}\Delta_L||_F \leq C_1\sqrt{\frac{p\log(NT)}{NT}}||\Delta_\beta||_F + \sqrt{3\bar{r}}\Big(\lambda + \frac{C_2\sqrt{N\vee T}}{NT}\Big)||P_{\Phi(u)}\Delta_L||_F$$

Hence, if $\lambda > \frac{C_2\sqrt{N\vee T}}{NT}$ and the first term on the right hand side is dominated by the second, the Frobenius norm of $P_{\Phi^\perp(u)}\Delta_L$ has at most the same order as $P_{\Phi(u)}\Delta_L$. Therefore, Δ_L is not "too far" from the space of L_0 in the Frobenius norm as well, which is useful to show the main results.

From now on, I set $\lambda = \frac{2C_2\sqrt{N\vee T}}{NT}$. Then the restricted set can be simplified as

$$\mathcal{R}_u \equiv \Big\{(\Delta_\beta, \Delta_L) : ||P_{\Phi^\perp(u)}\Delta_L||_* \leq \frac{C_1\sqrt{p\log(NT)(N\wedge T)}}{C_2}||\Delta_\beta||_F + 3||P_{\Phi(u)}\Delta_L||_*\Big\} \quad (3.2)$$

4 The Main Results

In this section I provide uniform rates for $\hat{\beta}(u)$ and $\hat{L}(u)$. I will also briefly discuss post-regularization procedures. Let us begin by introducing the following assumptions.

Assumption 2 (Conditional Density)**.** *The conditional densities satisfy $f_{V_{it}(u)|X_{it},F_t}(0) \geq \underline{f} > 0$ uniformly in $u\in\mathcal{U}$ almost surely. The derivative of the density is uniformly bounded in absolute value by $\bar{f}'$.*

Assumption 3 (Bounds on Magnitude)**.** $||L_0(u)||_\infty \leq \alpha$ *and* $\frac{\log(NT)}{\sqrt{N\wedge T}}||X_{it}||_F \leq \alpha^2$ *a.s..*

Assumption 4 (Smoothness)**.** *For any $u' \neq u \in \mathcal{U}$, there exist $\zeta_1, \zeta_2 > 0$ such that*

$$\begin{aligned} ||\beta_0(u') - \beta_0(u)||_F \leq& \zeta_1|u'-u|, \\ \frac{1}{\sqrt{NT}}||L_0(u') - L_0(u)||_F \leq& \zeta_2|u'-u| \text{ with high probability.} \end{aligned}$$

Assumption 5 (Identification)**.** *Denote* $(P_{\Phi(u)}\boldsymbol{X})_{it} = \big((P_{\Phi(u)}X_1)_{it}, ..., (P_{\Phi(u)}X_p)_{it}\big)'$ *and* $(P_{\Phi^\perp(u)}\boldsymbol{X})_{it} = \big((P_{\Phi^\perp(u)}X_1)_{it}, ..., (P_{\Phi^\perp(u)}X_p)_{it}\big)'$. *Conditional on* F_t*, assume the following holds uniformly in* $u \in \mathcal{U}$*:*

$$\mathbb{E}\sum_{it}\Big((P_{\Phi^\perp(u)}\boldsymbol{X})_{it}(P_{\Phi^\perp(u)}\boldsymbol{X})'_{it} - (\log(NT)\bar{r})(P_{\Phi(u)}\boldsymbol{X})_{it}(P_{\Phi(u)}\boldsymbol{X})'_{it}\Big) \text{ is positive definite.}$$

Assumptions 2 and 3 guarantee that the objective function can be bounded from below by a quadratic function. Assumption 2 is standard in quantile regression. However, with interactive fixed effects, a stronger assumption is often made requiring the conditional density to be bounded away from 0 on compact intervals around 0 (e.g. Ando and Bai (2019) and Chen, Dolado and Gonzalo (2019)). The stronger assumption is to overcome the theoretical difficulty caused by estimating the $N \times T$ matrix $L_0(u)$ without sparsity. In this paper, I develop a new argument under which the conditional density being bounded away from 0 at $V_{it}(u) = 0$ suffices. To apply the argument, I need to control the magnitude $L_{0,it}(u)$ and $||X_{it}||_F$ by Assumption 3. Since I allow $\alpha \to \infty$, these restrictions are mild in practice. I can further relax them by allowing the conditions only hold with sufficiently large probabilities. I maintain the stronger version for simplicity.

Assumption 4 is needed for uniformity. The first part is the same as in Belloni and Chernozhukov (2011). The second part imposes smoothness on the common component $L_0(\cdot)$. Note that the condition rules out the case where $r(u)$ changes on $\mathcal{U}$. To see this, suppose there exists u_0 in the interior of $\mathcal{U}$ such that $r(u) < r(u')$ for any $u < u_0 \leq u'$. Then $\frac{1}{NT}||L_0(u') - L_0(u)||_F$ is bounded away from 0 uniformly in $|u' - u|$. However, if $r(u)$ only has finite number of jump-points in $[0, 1]$, uniformity on the union of compact interior subsets between jump-points can be obtained.

Assumption 5 is used to bound the estimation errors $\hat{\Delta}_\beta(u)$ and $\hat{\Delta}_L(u)$ separately. Without it, I can only obtain the error bound for their weighted sum. This is a sufficient condition in our context to the "restricted strong convexity" condition; variants of the condition are widely assumed in the related literature (e.g. Negahban and Wainwright (2011, 2012), Agarwal, Negahban and Wainwright (2012), Negahban et al. (2012), Chernozhukov et al. (2019), Moon and Weidner (2019), etc.). Assumption 5 says the covariates need to be sufficiently far from the space of $L_0(u)$ uniformly in $u \in \mathcal{U}$. Otherwise, for instance, suppose for some j, $X_j = L_0(u)$, that is $X_j = P_{\Phi(u)}(X_j)$, then $\beta_{0,j}(u)$ is not identified due to perfect collinearity.

Under these assumptions, we have the main theorem of this paper.

Theorem 1. *Under Assumptions 1-5, there exists a constant* $C > 0$ *such that with high* *probability,*

$$\sup_{u\in\mathcal{U}} \ ||\hat{\Delta}_\beta(u)||_F^2 + \frac{1}{NT}||\hat{\Delta}_L(u)||_F^2 \leq C\frac{\alpha^4 \bar{f}'^4 \log(NT)}{\underline{f}^8}\Big(\frac{p\log(NT)}{NT} \vee \frac{\bar{r}}{N\wedge T}\Big) \tag{4.1}$$

Proof. See Appendix B. □

We see the rate of convergence depends on the rank of $L_0(u)$ (captured by $\bar{r}$), the number of regressors (p), and the magnitude of elements in $L_0(u)$ (α). Note that in the parentheses, $\frac{p}{NT}$ would be the rate of convergence of the standard quantile regression estimator if $L_0(u)$ were not present in the model. Meanwhile, $\frac{\bar{r}}{N\wedge T}$ is the minimax optimal rate of convergence for the low rank matrix estimator using nuclear norm regularization (see Agarwal, Negahban and Wainwright (2012) and Negahban and Wainwright (2012) for instance), and is identical to the mean regression (Athey et al. (2018), Moon and Weidner (2019) and Chernozhukov et al. (2019)). As the latter rate is much slower than the former (for fixed p), the rate of the coefficient estimate is much slower than its optimal rate.

From Theorem 1 and the definition of the restricted set, it is straightforward to have the following corollary.

Corollary 1. *Under Assumptions 1-5, there exists a constant $C' > 0$ such that with high probability,*

$$\sup_{u\in\mathcal{U}} \ \frac{1}{N\wedge T}||\hat{\Delta}_L(u)||_* \leq C'\frac{\alpha^2 \bar{f}'^2 \sqrt{\log(NT)}}{\underline{f}^4}\Big(\frac{\sqrt{p\log(NT)}}{N\wedge T} \vee \frac{\sqrt{\bar{r}(N\vee T)}}{N\wedge T}\Big) \tag{4.2}$$

To see why this is true, note that being in the restricted set, the rate in Theorem 1 implies that $||P_{\Phi^\perp(u)}\hat{\Delta}_L(u)||_*$ is bounded by $||P_{\Phi(u)}\hat{\Delta}_L(u)||_*$, which is further bounded by $\sqrt{3\bar{r}}||P_{\Phi(u)}\hat{\Delta}_L(u)||_F$. The low rankness of $L_0(u)$ plays a key role here because only then the error bound in the nuclear norm is of the same order of that in the Frobenius norm. Corollary 1 implies that if the panel data matrix is not too "tall" or "fat", i.e., the order of N and T are not too different, the average of the singular values of $L_0(u)$ can be uniformly estimated.

From Theorem 1 and Corollary 1, we can consistently estimate the rank of $L_0(u)$ by thresholding, similar to Moon and Weidner (2019) and Chernozhukov et al. (2019). Denote the singular values of $\hat{L}(u)$ by $\hat{\sigma}_1(u), \cdots, \hat{\sigma}_{N\wedge T}(u)$ in descending order. By Weyl's theorem and Theorem 1, $\max_k |\hat{\sigma}_k(u) - \sigma_k(u)|$ is bounded by $O(\sqrt{N\vee T})$ up to a multiplicative polynomial of $\log(NT)$. Since $\sigma_k(u) = 0$ for all $k > r(u)$, the singular values of $\hat{L}(u)$ are well separated if $F_t(u)$ are strong factors: they are either of the order of $(N\wedge T)\cdot\sqrt{N\vee T}$ ($k \leq r(u)$), or $\sqrt{N\vee T}$ ($k > r(u)$). Therefore, by setting a threshold of any order in between, $r(u)$ can be consistently estimated by simply counting the number of estimated singular values that are above the threshold.

To correct the bias that regularization brings in and achieve the optimal rate for $\hat{\beta}(u)$, I conjecture that one can follow similar post-regularization procedures proposed in Moon and Weidner (2019) and Chernozhukov et al. (2019) to obtain an asymptotically normal estimator for $\beta_0(u)$. Specifically, construct $\hat{F}_t$ and $\hat{\Lambda}_i$ by singular value decomposition with only the largest $\hat{r}(u)$ singular values kept. Then iteratively minimize the standard quantile regression objective function without penalty, similar to Ando and Bai (2019) but setting $\hat{\beta}(u)$ as the initial guess.

5 Monte Carlo Simulations

In this section, I illustrate the finite sample performance of our estimator using a simulation study. I consider the following data generating process adapted from Ando and Bai (2019):

$$Y_{it} = X_{it}'\beta_0(U_{it}) + \sum_{k=1}^{5} \mathbb{1}_k(U_{it}) F_{kt} \Lambda_{ki}(U_{it}) + V_{it}(U_{it})$$

where $U_{it} \sim \text{Unif}[0,1]$. X_{it} contains the following four regressors:

$$\begin{aligned} X_{1,it} &= W_{1,it} + 0.02F_{1,t}^2 + 0.02\zeta_{1,i}^2, \quad X_{2,it} = W_{2,it} \\ X_{3,it} &= W_{3,it} - 0.01F_{3,t}^2 + 0.02\zeta_{3,i}^2, \quad X_{4,it} = W_{4,it} - 0.01F_{4t}^2 + 0.03\zeta_{4,i}^2, \end{aligned}$$

where all $W_{k,it}$, $\zeta_{k,i}$ and the factors $F_{k,t}$ are independently drawn from Unif$[0,2]$. The factors loading $\Lambda_{k,i}(U_{it}) = \zeta_{k,i} + 0.1U_{it}$. For $\beta_0(U_{it})$, $\beta_{0,1}(U_{it}) = \beta_{0,3}(U_{it}) = \beta_{0,4}(U_{it}) = -1 + 0.1U_{it}$ and $\beta_{0,2}(U_{it}) = 1 + 0.1U_{it}$. Finally, the indicator function satisfies:

$$\mathbb{1}_k(\cdot) = 1, k = 1,2,3, \quad \mathbb{1}_4(u) = \mathbb{1}(0.3 < u \leq 0.7), \quad \mathbb{1}_5(u) = \mathbb{1}(u > 0.7),$$

that is, the first three factors always affect Y_{it}, while the fourth and the fifth factor only affect Y_{it} when U_{it} is between 0.3 and 0.7 and above 0.7, respectively. Finally, the error term $V_{it}(U_{it}) = G^{-1}(U_{it})$ where G is the cumulative distribution function of standard normal or student-t distribution with degree of freedom 8.

For implementation, I set $\lambda = \frac{1}{\sqrt{N}}$; the order is suggested by the theory. I then iteratively update $\beta(u)$ and $L(u)$ until convergence. I update $\beta(u)$ by pooled quantile regression subtracting $L(u)$ from Y. For $L(u)$, I adapt the *alternating directions* method proposed in Lin, Chen and Ma (2010) and Yuan and Yang (2013), which is also used in Candès et al.

Table 1: Average Bias, Variance and RMSE of $\hat{\beta}(u)$ and $\hat{\beta}^{pooled}$

	A. Normal Error				
	Bias^2		Variance		$\|\|\hat{\Delta}_L(u)\|\|_F^2/\|\|L_0(u)\|\|_F^2$
	Penalized	Pooled	Penalized	Pooled	Penalized
$u = 0.2$	$2.82 \cdot 10^{-4}$	$2.6 \cdot 10^{-3}$	$1.25 \cdot 10^{-4}$	$5.48 \cdot 10^{-4}$	0.07
$u = 0.5$	$7.40 \cdot 10^{-6}$	$2 \cdot 10^{-3}$	$1.46 \cdot 10^{-4}$	$5.72 \cdot 10^{-4}$	0.07
$u = 0.8$	$2.22 \cdot 10^{-5}$	$2.1 \cdot 10^{-3}$	$5.99 \cdot 10^{-4}$	$9.62 \cdot 10^{-4}$	0.06
	B. Student-t Error				
	Bias^2		Variance		$\|\|\hat{\Delta}_L(u)\|\|_F^2/\|\|L_0(u)\|\|_F^2$
	Penalized	Pooled	Penalized	Pooled	Penalized
$u = 0.2$	$2.82 \cdot 10^{-4}$	$2.6 \cdot 10^{-3}$	$1.45 \cdot 10^{-4}$	$5.40 \cdot 10^{-4}$	0.08
$u = 0.5$	$4.54 \cdot 10^{-6}$	$2 \cdot 10^{-3}$	$1.65 \cdot 10^{-4}$	$7.53 \cdot 10^{-4}$	0.07
$u = 0.8$	$2.06 \cdot 10^{-5}$	$2 \cdot 10^{-3}$	$7.17 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$	0.06

(2011). In our notation, the algorithm was originally designed for $u = 0.5$. I modify it to accommodate any other $u \in (0, 1)$.

Table 1 presents the results based on 100 simulation replications. The squared biases and the variances are the average of those of the four components in $\hat{\beta}(u)$. Column *penalized* contains the results using our estimator. Column *pooled* are results obtained from quantile regression ignoring the interactive fixed effects. Since the regularized regression is biased, column pooled provides a reference to compare the biases that the nuclear norm penalty introduces with the biases from endogeneity. From the results, we can see the squared biases ignoring the interactive fixed effects are 10-1000 times than our estimator. The performance of the estimated common component is relatively good too. The sum of squared estimation error is between 6% and 8% of the magnitude of the true common component.

6 Concluding Remarks

In this paper, I propose the nuclear norm regularized quantile regression for panel data models with interactive fixed effects. I derive the uniform rate of convergence for both the coefficients of the regressors and the common component. The rate for the common component is nearly optimal.

The results can be extended to models with heterogeneous effects. To see it, note that in the error bound I derive, the number of coefficients can be as large as $O(N \vee T)$, only increasing the order of the bound by $\log(NT)$. Therefore, almost the same rate could be obtained if p is fixed but $\beta_0(u)$ is i- or t-dependent. Then the number of coefficients is pN or pT, so that the rate for the coefficients becomes identical to that in Ando and Bai

(2019) up to a polynomial factor of $\log(NT)$. It is also possible to extend the results to high dimensional regressors, i.e., $p \geq NT$ with an extra ℓ_1 norm penalty in the objective function like Belloni and Chernozhukov (2011). Finally, as the penalty introduces bias, post-regularization estimation procedures is needed to obtain the asymptotic distribution and establish inference theory. This is left for future work.

Appendix A Proof of Lemma 1

Let $\nabla\boldsymbol{\rho}_u(\cdot)$ be the subgradiant of $\boldsymbol{\rho}_u$ evaluated at $\cdot$, then by definition $\nabla\boldsymbol{\rho}_u(V(u))$ is an $N\times T$ matrix of which the (i,t)-th element is

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

Conditional on $(X_1,...,X_p)$ and $L_0(u)$, $\nabla\boldsymbol{\rho}_u(V)$ is thus a random matrix with i.i.d. mean 0 entries bounded by $u\vee(1-u)$. We have the following lemma for $\nabla\boldsymbol{\rho}_u(V)$:

Lemma A1. *Under Assumption 1, there exist two constants $C_1, C_2 > 0$ such that the following hold with high probabilities:*

$$\sup_{u\in\mathcal{U}}\max_{1\le j\le p}\left|\left\langle\nabla\boldsymbol{\rho}_u(V(u)), X_j\right\rangle\right| \le C_1\sqrt{NT\log(NT)}, \tag{A.1}$$

$$\sup_{u\in\mathcal{U}}\max_{1\le j\le p}||\nabla\boldsymbol{\rho}_u(V(u))|| \le C_2\sqrt{N\vee T} \tag{A.2}$$

Proof. See Appendix C. □

In what follows, the derivation is under the event that equations (A.1) and (A.2) hold. By the definition of $(\hat\beta(u), \hat L(u))$, the following holds uniformly in $u\in\mathcal{U}$:

$$\frac{1}{NT}\Big[\boldsymbol{\rho}_u(Y-\sum_{j=1}^p X_j\hat\beta(u)-\hat L(u)) - \boldsymbol{\rho}_u(Y-\sum_{j=1}^p X_j\beta_0(u)-L_0(u))\Big] + \lambda\Big[||\hat L(u)||_* - ||L_0(u)||_*\Big] \le 0$$

$$\Longrightarrow \frac{1}{NT}\Big[\boldsymbol{\rho}_u\Big(V(u)-\sum_{j=1}^p X_j\hat\Delta_{\beta,j}(u)-\hat\Delta_L(u)\Big) - \boldsymbol{\rho}_u(V(u))\Big] + \lambda\Big[||\hat L(u)||_* - ||L_0(u)||_*\Big] \le 0$$

Let us first consider $\frac{1}{NT}\Big[\boldsymbol{\rho}_u\Big(V(u)-\sum_{j=1}^p X_j\hat\Delta_{\beta,j}(u)-\hat\Delta_L(u)\Big) - \boldsymbol{\rho}_u(V(u))\Big]$. By the definition of the subgradient,

$$\begin{aligned}
&\frac{1}{NT}\Big[\boldsymbol{\rho}_u\Big(V(u)-\sum_{j=1}^p X_j\hat\Delta_{\beta,j}(u)-\hat\Delta_L(u)\Big) - \boldsymbol{\rho}_u(V(u))\Big]\\
\ge& -\frac{1}{NT}\Big|\Big\langle\nabla\boldsymbol{\rho}_u(V(u)), \sum_{j=1}^p X_j\hat\Delta_{\beta,j}(u)+\hat\Delta_L(u)\Big\rangle\Big|\\
\ge& -\frac{1}{NT}||\hat\Delta_\beta(u)||_1\max_{1\le j\le p}\Big|\Big\langle\nabla\boldsymbol{\rho}_u(V(u)), X_j\Big\rangle\Big| - \frac{1}{NT}||\nabla\boldsymbol{\rho}_u(V(u))||\cdot||\hat\Delta_L(u)||_*\\
\ge& -C_1\sqrt{\frac{\log(NT)}{NT}}||\hat\Delta_\beta(u)||_1 - \frac{C_2\sqrt{N\vee T}}{NT}||\hat\Delta_L(u)||_*
\end{aligned}$$

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

Next, consider $\lambda\Big[||\hat{L}(u)||_* - ||L_0(u)||_*\Big]$. By construction, $P_{\Phi^\perp(u)}L_0(u) = 0$, so

$$
\begin{aligned}
||\hat{L}(u)||_* - ||L_0(u)||_* =& ||P_{\Phi(u)}L_0(u) + P_{\Phi(u)}\hat{\Delta}_L(u)||_* + ||P_{\Phi^\perp(u)}\hat{\Delta}_L(u)||_* - ||P_{\Phi(u)}L_0(u)||_* \\
\geq& ||P_{\Phi^\perp(u)}\hat{\Delta}_L(u)||_* - ||P_{\Phi(u)}\hat{\Delta}_L(u)||_*
\end{aligned}
$$

Combining the two pieces with rearrangement, we have the following uniformly in $u \in \mathcal{U}$:

$$
\Big(\lambda - \frac{C_2\sqrt{N \vee T}}{NT}\Big)||P_{\Phi^\perp(u)}\hat{\Delta}_L(u)||_* \leq C_1\sqrt{\frac{p\log(NT)}{NT}}||\hat{\Delta}_\beta(u)||_F + \Big(\lambda + \frac{C_2\sqrt{N \vee T}}{NT}\Big)||P_{\Phi(u)}\hat{\Delta}_L(u)||_*
$$

Appendix B Proof of Theorem 1

Throughout, I condition on the event that Assumption 4 holds and $(\hat{\Delta}_\beta(u), \hat{\Delta}_L(u)) \in \mathcal{R}_u$ uniformly in $u \in \mathcal{U}$. To prove the theorem, we want to show the following is impossible:

$$
\exists u \in \mathcal{U} : ||\hat{\Delta}_\beta(u)||_F^2 + \frac{1}{NT}||\hat{\Delta}_L(u)||_F^2 \geq t^2
$$

where t^2 is the right hand side of inequality (4.1).

Under Assumption 3 and by the definition of the estimator, $||\hat{\Delta}_L(u)||_\infty \leq 2\alpha$ a.s. by the triangular inequality. Let $\mathcal{D} \equiv \{(\Delta_\beta, \Delta_L) : \Delta_\beta \in \mathbb{R}^p, \Delta_L \in \mathbb{R}^{N\times T}, ||\Delta_L||_\infty \leq 2\alpha\}$. Then by convexity of the objective function and the definition of the estimator, the inequality above implies that

$$
\begin{aligned}
0 \geq \min_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta||_F^2 + \frac{1}{NT}||\Delta_L||_F^2 \geq t^2}} & \frac{1}{NT}\Big[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j\Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\Big] \\
& + \lambda\Big[||L_0(u) + \Delta_L||_* - ||L_0(u)||_*\Big]
\end{aligned}
$$

The proof consists of two main steps. In the first step (Lemmas 2 and 3), I bound the norm of $\sum_{j=1}^p X_j\Delta_{\beta,j} + \Delta_L$. In the second step (Lemma 4), I separate $\Delta_{\beta,j}$ and Δ_L by invoking Assumption 5. The first step is adapted from the proofs of Theorem 2 in Belloni and Chernozhukov (2011) and Theorem 3.2 in Chao, Härdle and Yuan (2019). A new theoretical challenge arises in the minoration step (Lemma 2) because of the high dimensional object Δ_L. I develop a new argument to handle it. Lemma 3 follows the two papers cited closely where I only highlight the differences that Δ_L brings into.

Since $\mathcal{R}_u$ is a cone and $\mathcal{D}$ is convex, by convexity of the objective function, the inequality sign in the constraint can be replaced with equality:

$$0 \geq \min_{\substack{(\Delta_\beta, \Delta_L) \in \mathcal{R}_u \cap \mathcal{D} \\ ||\Delta_\beta||_F^2 + \frac{1}{NT}||\Delta_L||_F^2 = t^2}} \frac{1}{NT}\Big[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j \Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\Big] + \lambda\Big[||L_0(u) + \Delta_L||_* - ||L_0||_*\Big]$$

Let us rewrite the minimand as follows:

$$\begin{aligned}
&\frac{1}{NT}\Big[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j \Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\Big] + \lambda\Big[||L_0(u) + \Delta_L||_* - ||L_0(u)||_*\Big] \\
=&\frac{1}{NT}\mathbb{E}\Big[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j \Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\Big] \\
&+ \frac{1}{\sqrt{NT}}\mathbb{G}_{NT}\Big(\rho_u\Big(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\Big) - \rho_u(V_{it}(u))\Big) \\
&+ \lambda\Big[||L_0(u) + \Delta_L||_* - ||L_0(u)||_*\Big]
\end{aligned}$$

The following lemmas provide bounds for the first two terms on the right hand side respectively.

Lemma 2 (Minoration)**.** *Under Assumptions 1-3, there exists a constant $c > 0$ such that the following holds uniformly in u.*

$$\frac{1}{NT}\mathbb{E}\Big[\boldsymbol{\rho}_u\Big(V(u) - \sum_{j=1}^p X_j \Delta_{\beta,j} - \Delta_L\Big) - \boldsymbol{\rho}_u(V(u))\Big] \geq \frac{c\underline{f}^4}{(\bar{\alpha}\bar{f}')^2 NT}\mathbb{E}||X_{it}'\Delta_\beta + \Delta_{L,it}||_F^2 \tag{B.1}$$

To prove Lemma 2, I need the following result which will be used to handle the high dimensional object Δ_L.

Lemma A2. *For all $w_1, w_2 \in \mathbb{R}$ and all $\kappa \in (0,1)$,*

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

Proof. See Appendix C. □

Proof of Lemma 2. By the Knight's identity (Knight, 1998), for any two scalars w_1 and w_2,

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

Let $w_1 = V_{it}(u)$ and $w_2 = X_{it}'\Delta_\beta + \Delta_{L,it}$, then by construction $\mathbb{E}(-w_2(u - \mathbb{1}_{w_1 \leq 0})) = 0$. Also, by Lemma A2, with probability 1,

$$\begin{aligned}&\int_0^{X_{it}'\Delta_\beta+\Delta_{L,it}}\Big(\mathbb{1}(V_{it}(u)\le s)-\mathbb{1}(V_{it}(u)\le 0)\Big)ds\\ \ge&\int_0^{\kappa(X_{it}'\Delta_\beta+\Delta_{L,it})}\Big(\mathbb{1}(V_{it}(u)\le s)-\mathbb{1}(V_{it}(u)\le 0)\Big)ds\end{aligned}$$

where $\kappa\equiv\frac{3\underline{f}^2}{8\alpha\bar{f}'}\in(0,1)$ for large N and T. Then by the law of iterated expectation and mean value theorem,

$$\begin{aligned}&\mathbb{E}\int_0^{X_{it}'\Delta_\beta+\Delta_{L,it}}\Big(\mathbb{1}(V_{it}(u)\le s)-\mathbb{1}(V_{it}(u)\le 0)\Big)ds\\ \ge&\mathbb{E}\int_0^{\kappa(X_{it}'\Delta_\beta+\Delta_{L,it})}\Big(\mathbb{1}(V_{it}(u)\le s)-\mathbb{1}(V_{it}(u)\le 0)\Big)ds\\ \ge&\mathbb{E}\int_0^{\kappa(X_{it}'\Delta_\beta+\Delta_{L,it})}\Big(F_{V_{it}(u)|X_{it},L_{0,it}}(s)-F_{V_{it}(u)|X_{it},L_{0,it}}(0))\Big)ds\\ =&\mathbb{E}\int_0^{\kappa(X_{it}'\Delta_\beta+\Delta_{L,it})}\Big(sf_{V_{it}(u)|X_{it},L_{0,it}}(0)+\frac{s^2}{2}f'_{V_{it}(u)|X_{it},L_{0,it}}(\tilde{s}))\Big)ds\\ \ge&\frac{\kappa^2\underline{f}^2}{4}\mathbb{E}\Big(X_{it}'\Delta_\beta+\Delta_{L,it}\Big)^2+\mathbb{E}\Big[\frac{\kappa^2\underline{f}^2}{4}\Big(X_{it}'\Delta_\beta+\Delta_{L,it}\Big)^2\Big(1-|\frac{2\kappa\bar{f}'}{3\underline{f}^2}(X_{it}'\Delta_\beta+\Delta_{L,it}|\Big)\Big]\\ \ge&\frac{\kappa^2\underline{f}^2}{4}\mathbb{E}\Big(X_{it}'\Delta_\beta+\Delta_{L,it}\Big)^2\end{aligned}$$

where the third line is from the law of iterated expectation. The last inequality holds because under the choice of κ and t, $1-|\frac{2\kappa\bar{f}'}{3\underline{f}^2}(X_{it}'\Delta_\beta+\Delta_{L,it}|>0$ a.s. $\square$

Remark 1. As mentioned, Δ_L introduces new difficulties for minoration. Specifically, $||\Delta_L||_F^2$ can be greater than $\sum_{i,t}|\Delta_{L,it}|^3$ even in the restricted set $\mathcal{R}_u$. As a consequence, standard argument fails because after Taylor expansion the higher order term may be greater than the leading term, resulting in a negative lower bound for the expectation under investigation. I overcome this difficulty by exploiting the monotonicity in the residual integral and impose condition that directly restricts the tail behavior of $X_{j,it}$ and $L_{0,it}$.

Lemma 3 (Bound on the Empirical Process). *Under Assumptions 1, 2 and 4, there exists a constant $C_0>0$ such that with high probability*

$$\begin{aligned}&\sup_{\substack{u\in\mathcal{U}\\(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D}\\||\Delta_\beta||_F^2+\frac{1}{NT}||\Delta_L||_F^2=t^2}}\Big|\mathbb{G}_{NT}\Big(\rho_u\Big(V_{it}(u)-X_{it}'\Delta_\beta-\Delta_{L,it}\Big)-\rho_u(V_{it}(u))\Big)\Big|\\ \le&C_0\Big(\sqrt{p\log(NT)}+\sqrt{N\vee T}\sqrt{\bar{r}}\Big)\sqrt{\log(NT)}t\end{aligned}$$

Proof of Lemma 3. Note that the check function is a contraction. Hence, similar to Belloni and Chernozhukov (2011) and Chao, Härdle and Yuan (2019), there exists $C > 0$ such that

$$
\begin{aligned}
&\mathrm{Var}\Big(\mathbb{G}_{NT}\Big(\rho_u\Big(V_{it}(u) - X_{it}'\Delta_\beta - \Delta_{L,it}\Big) - \rho_u(V_{it}(u))\Big)\Big) \\
\leq& \frac{1}{NT}\sum_{i,t}\mathbb{E}(X_{it}'\Delta_\beta + \Delta_{L,it})^2 \\
\leq& \frac{2}{NT}\sum_{i,t}\mathbb{E}(X_{it}'\Delta_\beta)^2 + \frac{2}{NT}||\Delta_L||_F^2 \\
\leq& C\Big(||\Delta_\beta||_F^2 + \frac{1}{NT}||\Delta_L||_F^2\Big) \\
\leq& Ct^2
\end{aligned}
$$

where the third inequality follows from $\mathbb{E}(X_{it}'\Delta_\beta)^2 = \Delta_\beta'\mathbb{E}(X_{it}X_{it}')\Delta_\beta \leq C'\sigma_{\max}^2||\Delta_\beta||_F^2$ for some $C' > 0$, where $\sigma_{\max}^2$ is the largest eigenvalue of $.\mathbb{E}(X_{it}X_{it}')$.

Let $\mathcal{A}(t)$ denote the empirical process under investigation. Then by Lemma 2.3.7 in van der Vaart and Wellner (1996), let $s \geq 4t$, we have

$$
\mathbb{P}(\mathcal{A}(t) > s) \leq C''\mathbb{P}(\mathcal{A}^0(t) > \frac{s}{4})
$$

for some $C'' > 0$ where $\mathcal{A}^0(t)$ is the symmetrized version of $\mathcal{A}(t)$ by replacing $\mathbb{G}_{NT}$ with the symmetrized version $\mathbb{G}_{NT}^0$.

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

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

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

$$
\mathcal{B}_1^0(t) \equiv \sup_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta||_F^2\leq t^2}} \Big|\mathbb{G}^0\Big(X_{it}'\Delta_\beta\Big)\Big|,
$$

$$
\mathcal{B}_2^0(t) \equiv \sup_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta||_F^2+\frac{1}{NT}||\Delta_L||_F^2\leq t^2}} \Big|\mathbb{G}^0\Big(\Delta_{L,it}\Big)\Big|,
$$

and

$$
\mathcal{C}^0(t) \equiv \sup_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta||_F^2+\frac{1}{NT}||\Delta_L||_F^2\leq t^2}} \Big|\mathbb{G}^0\Big(\delta_{it}\Big(X_{it}'\Delta_\beta + \Delta_{L,it}, u\Big)\Big)\Big|,
$$

then $\mathcal{A}^0(t) \leq \mathcal{B}_1^0(t) + \mathcal{B}_2^0(t) + \mathcal{C}^0(t)$. Next I bound $\mathcal{B}_1^0(t)$, $\mathcal{B}_2^0(t)$ and $\mathcal{C}^0(t)$ respectively.

Bound on $\mathcal{B}_1^0(t)$. Since $\mathcal{B}_1^0(t)$ does not contain Δ_L, the bound is identical to that in Belloni and Chernozhukov (2011), i.e., $\mathcal{B}_1^0(t) \leq C_1\sqrt{p\log(NT)}t$ with high probability.

Bound on $\mathcal{B}_2^0(t)$. Let $(\varepsilon_{it} : i \in \{1,..,N\}, t \in \{1,...,T\})$ be the Rademacher sequence in the symmetrized process. Let $\boldsymbol{\varepsilon}$ be the $N \times T$ matrix containing all the elements in the sequence. Then

$$
\begin{aligned}
\mathcal{B}_2^0(t) \leq & \frac{1}{\sqrt{NT}} \sup_{\substack{u\in\mathcal{U} \\ (\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq t^2}} |\sum_{it}\varepsilon_{it}\Delta_{L,it}| \\
= & \frac{1}{\sqrt{NT}} \sup_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq t^2}} |\langle\boldsymbol{\varepsilon}, \Delta_L\rangle| \\
\leq & \frac{1}{\sqrt{NT}}||\boldsymbol{\varepsilon}|| \cdot \sup_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq t^2}} ||\Delta_L||_* \\
\leq & \frac{1}{\sqrt{NT}}||\boldsymbol{\varepsilon}|| \cdot \sup_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq t^2}} (||P_{\Phi(u)}\Delta_L||_* + ||P_{\Phi^\perp(u)}\Delta_L||_*) \\
\leq & \frac{1}{\sqrt{NT}}||\boldsymbol{\varepsilon}|| \cdot \sup_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D} \\ ||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq t^2}} \left(C_2'\sqrt{p(N\wedge T)\log(NT)}||\Delta_\beta||_F + 4C_3||\Delta_L||_*\right) \\
\leq & \frac{1}{\sqrt{NT}}||\boldsymbol{\varepsilon}|| \cdot \left(C_2'\sqrt{p(N\wedge T)\log(NT)} + C_4\sqrt{NT\bar{r}}\right)t
\end{aligned}
$$

where the second to the last inequality is from the definition of the restricted set $\mathcal{R}_u$. Finally, since elements in $\boldsymbol{\varepsilon}$ are i.i.d. mean 0 and uniformly bounded in magnitude by 1, the operator norm is bounded by $C_5\sqrt{N \vee T}$ with high probability (Corollary 2.3.5 in Tao (2012)). Therefore,

$$
\mathcal{B}_2^0(t) \leq \left(C_6\sqrt{p\log(NT)} + C_7\sqrt{N\vee T}\sqrt{\bar{r}}\right)t
$$

with high probability.

Bound on $\mathcal{C}^0(t)$. The smoothness Assumption 3 allows us to follow similar ϵ-net argument in Belloni and Chernozhukov (2011). Since there is a new term Δ_L in our case, I write down the proof for completeness. Let $\mathcal{U}_l = \{u_1, ..., u_l\}$ be an ϵ-net in $\mathcal{U}$ where $\epsilon \leq t$. By the triangular inequality, we have

$$
\begin{aligned}
\mathcal{C}^0(t) \leq & \sup_{\substack{u\in\mathcal{U},|u-u_l|<\epsilon,u_l\in\mathcal{U}_l\\(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D}\\||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq t^2}} \left|\mathbb{G}_{NT}^0\Big(\delta_{it}[X_{it}'(\Delta_\beta+\beta_0(u)-\beta_0(u_l))+\Delta_L+L_0(u)-L_0(u_l)]),u_l)\Big)\right| \\
& + \sup_{\substack{u\in\mathcal{U},|u-u_l|<\epsilon,u_l\in\mathcal{U}_l\\(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D}\\||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq t^2}} \left|\mathbb{G}_{NT}^0\Big(\delta_{it}[X_{it}'(\beta_0(u)-\beta_0(u_l))+L_0(u)-L_0(u_l),u_l]\right| \\
& \leq 2\cdot \sup_{\substack{u_l\in\mathcal{U}_l\\(\Delta_\beta,\Delta_L)\in\bar{\mathcal{R}}\cap\bar{\mathcal{D}}\\||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq(\zeta_1^2+\zeta_2^2+1)t^2}} \left|\mathbb{G}_{NT}^0\Big(\delta_{it}(X_{it}'\Delta_\beta+\Delta_L,u_l)\Big)\right|
\end{aligned}
$$

where the last inequality follows from Assumption 4 that $\sup_{|u-u_l|<\epsilon}||\beta_0(u)-\beta_0(u_l)||_F\leq\zeta_1\epsilon$ and $\frac{1}{\sqrt{NT}}\sup_{|u-u_l|<\epsilon}||L_0(u)-L_0(u_l)||_F\leq\zeta_2\epsilon$, and thus under the choice of ϵ, we can treat $\Delta_\beta+\beta_0(u)-\beta_0(u_l)$ and $\beta_0(u)-\beta_0(u_l)$ as new Δ_β. Similarly, I treat $\Delta_L+L_0(u)-L_0(u_l)$ and $L_0(u)-L_0(u_l)$ as new Δ_L by adapting $\mathcal{R}_u$ and $\mathcal{D}$. $\bar{\mathcal{D}}$ contains all $||\Delta_L||_\infty\leq 4\alpha$. For $\bar{\mathcal{R}}$, note that $L_0(u_l)$ may not be in the space of $L_0(u)$, so it may be the case that $L_0(u)-L_0(u_l)\notin\mathcal{R}_u$. However, since

$$
\operatorname{rank}(L_0(u)-L_0(u_l))\leq r(u)+r(u_l)\leq 2\bar{r},
$$

$||L_0(u)-L_0(u_l)||_*\leq\sqrt{2\bar{r}}||L_0(u)-L_0(u_l)||_F\leq\sqrt{2\bar{r}}\zeta_2\sqrt{NT}t$ by Assumption 4. From the derivation for the bound on $\mathcal{B}_2^0(t)$, set $\bar{\mathcal{R}}=\{\Delta_L:||L||_*\leq(C_2'\sqrt{pNT\log(NT)}+(C_4+\zeta_2)\sqrt{NT}\sqrt{2\bar{r}})t\}$.

Now by Markov inequality,

$$
\begin{aligned}
&\mathbb{P}\Big(\mathcal{C}^0(t)\geq\Big(C_8\sqrt{p\log(NT)}+C_9\sqrt{N\vee T}\sqrt{\bar{r}}\Big)\sqrt{\log(NT)}t\Big)\\
&\leq\min_{\tau\geq 0}e^{-\tau(C_8\sqrt{p\log(NT)}+C_9\sqrt{N\vee T}\sqrt{\bar{r}})\log(NT)t}\mathbb{E}[e^{\tau\mathcal{C}^0(t)}]
\end{aligned}
$$

By Theorem 4.12 of Ledoux and Talagrand (1991), the contractivity of $\delta_{it}(\cdot)$ implies

$$
\begin{aligned}
\mathbb{E}[e^{\tau\mathcal{C}^0(t)}]\leq&(1/\epsilon)\max_{u_l\in\mathcal{U}_l}\mathbb{E}\Big[\exp\Big(2\tau\sup_{\substack{(\Delta_\beta,\Delta_L)\in\bar{\mathcal{R}}\cap\bar{\mathcal{D}}\\||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq(1+\zeta^2+\zeta^2)t^2}}\left|\mathbb{G}_{NT}^0\Big(\delta_{it}(X_{it}'\Delta_\beta+\Delta_{L,it},u_l)\Big)\right|\Big)\Big]\\
\leq&(1/\epsilon)\mathbb{E}\Big[\exp\Big(4\tau\sup_{\substack{(\Delta_\beta,\Delta_L)\in\bar{\mathcal{R}}\cap\bar{\mathcal{D}}\\||\Delta_\beta(u)||_F^2+\frac{1}{NT}||\Delta_L(y)||_F^2\leq(1+\zeta^2+\zeta^2)t^2}}\left|\mathbb{G}_{NT}^0\Big(X_{it}'\Delta_\beta+\Delta_{L,it}\Big)\right|\Big)\Big]\\
\leq&(1/\epsilon)\mathbb{E}\Big[\exp\Big(4\tau\sup_{\substack{(\Delta_\beta,\Delta_L)\in\bar{\mathcal{R}}\cap\bar{\mathcal{D}}\\||\Delta_\beta(u)||_F^2\leq(1+\zeta^2+\zeta^2)t^2}}\left|\mathbb{G}_{NT}^0\Big(X_{it}'\Delta_\beta\Big)\right|+\sup_{\substack{(\Delta_\beta,\Delta_L)\in\bar{\mathcal{R}}\cap\bar{\mathcal{D}}\\||\frac{1}{NT}||\Delta_L(y)||_F^2\leq(1+\zeta^2+\zeta^2)t^2}}\left|\mathbb{G}_{NT}^0\Delta_{L,it}\Big)\right|\Big)\Big]
\end{aligned}
$$

Following exactly the same argument for $\mathcal{B}_1^0$ and $\mathcal{B}_2^0$, we obtain

$$\mathcal{C}^0(t) \leq \Big(C_8\sqrt{p\log(NT)} + C_9\sqrt{N \vee T}\sqrt{\bar{r}}\Big)\sqrt{\log(NT)}t$$

with high probability. $\square$

Finally consider the difference in the penalty. From the derivation of the bound on $\mathcal{B}_2^0$, we have

$$\lambda\Big|||L_0(u)+\Delta_L||_* - ||L_0||_*\Big| \leq \lambda||\Delta_L||_* \leq \lambda\Big(C_2'\sqrt{p(N\wedge T)\log(NT)} + C_4\sqrt{NT\bar{r}}\Big)t$$

By the choice of λ, the right hand side is $(C_{10}\frac{\sqrt{p}\log(NT)}{\sqrt{NT}} + C_{11}\frac{\bar{r}}{\sqrt{N\wedge T}})t$ for some $C_{10}, C_{11} > 0$.

Combining all the pieces together, we have

$$\begin{aligned}
&\min_{\substack{(\Delta_\beta,\Delta_L)\in\mathcal{R}_u\cap\mathcal{D}\\ ||\Delta_\beta||_F^2+\frac{1}{NT}||\Delta_L||_F^2=t^2}} \frac{1}{NT}\Big[\boldsymbol{\rho}_u\Big(V(u)-\sum_{j=1}^p X_j\Delta_{\beta,j}-\Delta_L\Big)-\boldsymbol{\rho}_u(V(u))\Big] + \lambda\Big[||L_0(u)+\Delta_L||_* - ||L_0(u)||_*\Big]\\
\geq&\frac{cf^4}{(\alpha\bar{f}')^2NT}\mathbb{E}||\sum_{j=1}^p X\Delta_{\beta,j}+\Delta_L||_F^2 - \bar{C}\Big(\frac{\sqrt{p}\log(NT)}{\sqrt{NT}}+\frac{\sqrt{\bar{r}}}{\sqrt{N\wedge T}}\Big)\log(NT)t
\end{aligned}$$

uniformly in u over $\mathcal{U}$ for some $\bar{C} > 0$.

To obtain the error bound in the claim of Theorem 1, I need to separate the two terms in the the expectation. This is guaranteed by Assumption 5.

Lemma 4 (Separation)**.** *Under Assumption 1 and 5, for any $(\Delta_\beta, \Delta_L) \in \mathbb{R}_u$ such that $||\Delta_\beta||_F^2 + \frac{1}{NT}||\Delta_L||_F^2 = t^2$, there exists $C > 0$ such that*

$$\frac{1}{NT}\mathbb{E}||\sum_{j=1}^p X_j\Delta_{\beta,j}+\Delta_L||_F^2 \geq Ct^2$$

Proof of Lemma 4. I prove the result conditional on F_t, and then by the law of iterated expectation the desired result is obtained. In the following, all the expectations are implicitly conditional on F_t. $P_{\Phi(u)}$ and $P_{\Phi^\perp u}$ are then nonrandom operators. Given $||\Delta_\beta||_F^2 + \frac{1}{NT}||\Delta_L||_F^2 = t^2$, let $||\Delta_\beta||_F^2 = \gamma t^2$. Then $\frac{1}{NT}||\Delta_L||_F^2 = (1-\gamma)t^2$. I am to express the expectation in the claim as a function of γ and show it is bounded away from a fixed fraction of t^2 uniformly in $\gamma \in [0,1]$ and F_t under Assumptions 1 and 5.

First consider $\mathbb{E}||\sum_{j=1}^p X_j\Delta_{\beta,j}||_F^2$. Let $\sigma_{\min}^2$ denote the smallest eigenvalues of $\mathbb{E}(X_{it}X_{it}')$.

By Assumption 1, $\sigma^2_{\min} > 0$. Recall that $\sigma^2_{\max}$ denote its largest eigenvalues. By i.i.d.,

$$\mathbb{E}||\sum_{j=1}^{p} X_j\Delta_{\beta,j}||_F^2 = NT\Delta_\beta'\mathbb{E}(X_{it}X_{it}')\Delta_\beta \in [NT\sigma^2_{\min}\gamma t^2, NT\sigma^2_{\max}\gamma t^2]$$

Hence, by Assumption 1, there exists a positive c_0 bounded away from 0 uniformly in γ such that $\mathbb{E}||\sum_{j=1}^{p} X_j\Delta_{\beta,j}||_F^2 = NTc_0\gamma t^2$.

Meanwhile, by the Pythagoras theorem,

$$\begin{aligned}\mathbb{E}||\sum_{j=1}^{p} X_j\Delta_{\beta,j}||_F^2 =& \mathbb{E}||P_{\Phi(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 + \mathbb{E}||P_{\Phi^\perp(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 \\ =& \Delta_\beta'\sum_{i,t}\mathbb{E}\Big((P_{\Phi(u)}\boldsymbol{X})_{it}(P_{\Phi(u)}\boldsymbol{X})_{it}'\Big)\Delta_\beta + \Delta_\beta'\sum_{i,t}\mathbb{E}\Big((P_{\Phi^\perp(u)}\boldsymbol{X})_{it}(P_{\Phi^\perp(u)}\boldsymbol{X})_{it}'\Big)\Delta_\beta\end{aligned}$$

Let $c_1\Delta_\beta'\sum_{i,t}\mathbb{E}\Big((P_{\Phi(u)}\boldsymbol{X})_{it}(P_{\Phi(u)}\boldsymbol{X})_{it}'\Big)\Delta_\beta = \Delta_\beta'\sum_{i,t}\mathbb{E}\Big((P_{\Phi^\perp(u)}\boldsymbol{X})_{it}(P_{\Phi^\perp(u)}\boldsymbol{X})_{it}'\Big)\Delta_\beta$.

Next consider Δ_L. Since it is in the restricted set, by the derivation in Section 3,

$$||P_{\Phi^\perp(u)}\Delta_L||_F \leq \frac{C_1\sqrt{p\log(NT)(N\wedge T)}}{C_2}\sqrt{\gamma}t + 3\sqrt{3\bar{r}}||P_{\Phi(u)}\Delta_L||_F$$

Let $c_2||P_{\Phi(u)}\Delta_L||_F^2 = ||P_{\Phi^\perp(u)}\Delta_L||_F^2$, then by $||P_{\Phi(u)}\Delta_L||_F^2 + ||P_{\Phi^\perp(u)}\Delta_L||_F^2 = NT(1-\gamma)t^2$, we have

$$\sqrt{\frac{c_2}{c_2+1}}\sqrt{1-\gamma} \leq \frac{C_1\sqrt{\log(NT)}}{C_2\sqrt{N\vee T}}\sqrt{p\gamma} + 3\sqrt{3\bar{r}}\sqrt{\frac{1-\gamma}{c_2+1}} \tag{A.3}$$

Now let us consider the expectation under investigation.

$$
\begin{aligned}
&\mathbb{E}||\sum_{j=1}^{p} X_j\Delta_{\beta,j} + \Delta_L||_F^2\\
=&\mathbb{E}||P_{\Phi(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 + ||P_{\Phi(u)}\Delta_L||_F^2 + 2\mathbb{E}\langle\sum_{j=1}^{p} P_{\Phi(u)}X_j\Delta_{\beta,j}, P_{\Phi(u)}\Delta_L\rangle\\
&+\mathbb{E}||P_{\Phi^\perp(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 + ||P_{\Phi^\perp(u)}\Delta_L||_F^2 + 2\mathbb{E}\langle\sum_{j=1}^{p} P_{\Phi^\perp(u)}X_j\Delta_{\beta,j}, P_{\Phi^\perp(u)}\Delta_L\rangle\\
\geq&\mathbb{E}||P_{\Phi(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 + ||P_{\Phi(u)}\Delta_L||_F^2 - 2\mathbb{E}||\sum_{j=1}^{p} P_{\Phi(u)}X_j\Delta_{\beta,j}||_F||P_{\Phi(u)}\Delta_L||_F\\
&+\mathbb{E}||P_{\Phi^\perp(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 + ||P_{\Phi^\perp(u)}\Delta_L||_F^2 - 2\mathbb{E}||\sum_{j=1}^{p} P_{\Phi^\perp(u)}X_j\Delta_{\beta,j}||_F||P_{\Phi^\perp(u)}\Delta_L||_F\\
\geq&\mathbb{E}||P_{\Phi(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 + ||P_{\Phi(u)}\Delta_L||_F^2 - 2\sqrt{\mathbb{E}||\sum_{j=1}^{p} P_{\Phi(u)}X_j\Delta_{\beta,j}||_F^2}||P_{\Phi(u)}\Delta_L||_F\\
&+\mathbb{E}||P_{\Phi^\perp(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2 + ||P_{\Phi^\perp(u)}\Delta_L||_F^2 - 2\sqrt{\mathbb{E}||\sum_{j=1}^{p} P_{\Phi^\perp(u)}X_j\Delta_{\beta,j}||_F^2}||P_{\Phi^\perp(u)}\Delta_L||_F\\
=&\Big(\sqrt{\mathbb{E}||P_{\Phi(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2} - ||P_{\Phi(u)}\Delta_L||_F\Big)^2 + \Big(\sqrt{\mathbb{E}||P_{\Phi^\perp(u)}(\sum_{j=1}^{p} X_j\Delta_{\beta,j})||_F^2} - ||P_{\Phi^\perp(u)}\Delta_L||_F\Big)^2\\
=&\Big[\Big(\sqrt{\frac{c_0\gamma}{1+c_1}} - \sqrt{\frac{1-\gamma}{1+c_2}}\Big)^2 + \Big(\sqrt{\frac{c_0c_1\gamma}{1+c_1}} - \sqrt{\frac{c_2(1-\gamma)}{1+c_2}}\Big)^2\Big]NTt^2
\end{aligned}
$$

where the last inequality follows from convexity of $||\cdot||_F^2$ and Jensen's inequality.

Now let us show the term $\Big[\Big(\sqrt{\frac{c_0\gamma}{1+c_1}} - \sqrt{\frac{1-\gamma}{1+c_2}}\Big)^2 + \Big(\sqrt{\frac{c_0c_1\gamma}{1+c_1}} - \sqrt{\frac{c_2(1-\gamma)}{1+c_2}}\Big)^2\Big]$ has a positive minimum.

First, fom equation (A.3), if $\frac{C_1\sqrt{\log(NT)}}{C_2\sqrt{N\vee T}}\sqrt{p\gamma}$ dominates $3\sqrt{3\bar{r}}\sqrt{\frac{1-\gamma}{c_2+1}}$ in order, then $(1-\gamma) = o(1)$, and both $\sqrt{\frac{1-\gamma}{c_2+1}}$ and $\sqrt{\frac{c_2}{c_2+1}}\sqrt{1-\gamma}$ are dominated by $\sqrt{p\gamma}$. Therefore, as long as there exists some positive constant ϖ such that $c_1 > \varpi$, the second term in the bracket is greater than $\frac{c_0\varpi}{2(1+\varpi)}$ for large enough N and T.

Now consider the case that $3\sqrt{3\bar{r}}\sqrt{\frac{1-\gamma}{c_2+1}}$ weakly dominates $\frac{C_1\sqrt{\log(NT)}}{C_2\sqrt{N\vee T}}\sqrt{p\gamma}$. Then for large enough N and T, $c_2 \leq C_0\bar{r}$ for some $C_0 > 0$. We have

$$
\begin{aligned}
&\Big(\sqrt{\frac{c_0\gamma}{1+c_1}} - \sqrt{\frac{1-\gamma}{1+c_2}}\Big)^2 + \Big(\sqrt{\frac{c_0c_1\gamma}{1+c_1}} - \sqrt{\frac{c_2(1-\gamma)}{1+c_2}}\Big)^2\\
=&c_0\gamma + (1-\gamma) - 2\Big(\sqrt{\frac{1}{(1+c_1)(1+c_2)}} + \sqrt{\frac{c_1c_2}{(1+c_1)(1+c_2)}}\Big)\sqrt{c_0\gamma(1-\gamma)}
\end{aligned}
$$

It can be verified that the right hand side is bounded away from 0 if there exists a constant $\eta > 0$ such that

$$\sqrt{\frac{1}{(1+c_1)(1+c_2)}} + \sqrt{\frac{c_1 c_2}{(1+c_1)(1+c_2)}} < 1 - \eta$$

The inequality holds if $c_1 - c_2 > \eta'$ for some $\eta' > 0$. This is the case if $c_1 > C_0' \bar{r}$ for some $C_0' > C_0$. By the definition of c_1, it is equivalent to

$$\Delta_\beta' \sum_{i,t} \mathbb{E}\Big((P_{\Phi^\perp(u)} \boldsymbol{X})_{it} (P_{\Phi^\perp(u)} \boldsymbol{X})_{it}' - C_0' \bar{r} (P_{\Phi(u)} \boldsymbol{X})_{it} (P_{\Phi(u)} \boldsymbol{X})_{it}' \Big) \Delta_\beta > 0,$$

which is guaranteed by Assumption 5. $\square$

This completes the proof.

Appendix C Proof of Lemmas A1 and A2

Proof of Lemma A1. Under Assumption 1, there exists a constant $C > 0$ such that $\max_{1 \le j \le p} ||X_j||_F^2 \le C\sqrt{NT}$ with high probability. In what follows, all probabilities and expectations are implicitly taken conditional on this event, $\{X_j\}$ and $L_0(u)$.

Proof of Equation (A.1).

Let $\mathcal{U}_K = (u_1, u_2, ..., u_K)$ be an ε-net in $\mathcal{U}$. Let $\varepsilon = \frac{1}{\sqrt{NT}}$. Then

$$\sup_{u \in \mathcal{U}} \Big| \Big\langle \nabla \boldsymbol{\rho}_u(V(u)), X_j \Big\rangle \Big| \le \max_{u_k \in \mathcal{U}_K} \Big| \Big\langle \nabla \boldsymbol{\rho}_{u_k}(V(u_k)), X_j \Big\rangle \Big| + \sup_{|u-u_k| \le \varepsilon, u_k \in \mathcal{U}_K} \Big| \Big\langle \nabla \boldsymbol{\rho}_u(V(u)) - \nabla \boldsymbol{\rho}_{u_k}(V(u_k)), X_j \Big\rangle \Big|$$

For the first term, since the length of $\mathcal{U}_K$ is no greater than 1,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{u_k \in \mathcal{U}_K} \Big| \Big\langle \nabla \boldsymbol{\rho}_{u_k}(V(u_k)), X_j \Big\rangle \Big| \ge C_1 \sqrt{NT \log(NT)} \Big) \\
\le & \frac{1}{\varepsilon} \max_{u_k \in \mathcal{U}_K} \mathbb{P}\Big(\Big| \Big\langle \nabla \boldsymbol{\rho}_{u_k}(V(u_k)), X_j \Big\rangle \Big| \ge C_1 \sqrt{NT \log(NT)} \Big) \\
\le & \frac{2}{\varepsilon} \exp\Big(- \frac{2C_1^2 \log(NT) NT}{||X_j||_F^2} \Big) \\
\le & \frac{2}{\varepsilon} \exp\Big(- \frac{2C_1^2 \log(NT) NT}{CNT} \Big) \\
\le & \frac{C_1'}{\sqrt{NT}}
\end{aligned}$$

where the second line is by Hoeffding's inequality.

For the second term, by definition, the (i,t)-th element in $\nabla \boldsymbol{\rho}_u(V(u)) - \nabla \boldsymbol{\rho}_{u_k}(V(u_k))$ is

$$\begin{aligned} & u\mathbb{1}_{V_{it}(u)>0} + (u-1)\mathbb{1}_{V_{it}(u)<0} - [u_k\mathbb{1}_{V_{it}(u_k)>0} + (u_k-1)\mathbb{1}_{V_{it}(u_k)<0}] \\ =&(u-u_k) + u_k(\mathbb{1}_{V_{it}(u)>0} - \mathbb{1}_{V_{it}(u_k)>0}) + (u_k-1)(\mathbb{1}_{V_{it}(u)<0} - \mathbb{1}_{V_{it}(u_k)<0}) \end{aligned}$$

The first term forms a constant matrix M_1 such that $||M_1||_F = \sqrt{NT}(u-u_k) \leq \varepsilon\sqrt{NT}$. Therefore, by the Cauchy-Schwartz inequality,

$$\left|\langle M_1, X_j\rangle\right| \leq ||M_1||_F||X_j||_F \leq CNT\varepsilon < C\sqrt{NT\log(NT)}$$

For the remaining terms, let $\xi_{it}(u,u_k) = u_k(\mathbb{1}_{V_{it}(u)>0} - \mathbb{1}_{V_{it}(u_k)>0}) + (u_k-1)(\mathbb{1}_{V_{it}(u)<0} - \mathbb{1}_{V_{it}(u_k)<0}) = \mathbb{1}_{V_{it}(u_k)<0} - \mathbb{1}_{V_{it}(u)<0} = \mathbb{1}_{U_{it}<u_k} - \mathbb{1}_{U_{it}<u}$ where the last equality is from the definition of $V_{it}(u)$. Therefore, if $u < u_k$, $0 \leq \xi_{it}(u,u_k) \leq \xi_{it}^{(1)}(u_k) = \mathbb{1}_{U_{it}<u_k} - \mathbb{1}_{U_{it}<u_{k-1}} \leq 1$. If $u \geq u_k$, $0 \geq \xi_{it}(u,u_k) \geq \xi_{it}^{(2)}(u_k) = \mathbb{1}_{U_{it}<u_k} - \mathbb{1}_{U_{it}<u_{k+1}} \geq -1$. Let $M_2^o = (\xi_{it}(u,u_k))_{1\leq i\leq N, 1\leq j\leq T}$, $M_2^{(1)} \equiv (\xi_{it}^{(1)}(u_k))_{1\leq i\leq N, 1\leq j\leq T}$, and $M_2^{(2)} \equiv (\xi_{it}^{(2)}(u_k))_{1\leq i\leq N, 1\leq j\leq T}$, we have

$$\begin{aligned} \sup_{|u-u_k|\leq\varepsilon, u_k\in\mathcal{U}_k} \left|\langle M_2^o, X_j\rangle\right| &\leq \sup_{u_k-\varepsilon\leq u\leq u_k, u_k\in\mathcal{U}_k} \left|\langle M_2^o, X_j\rangle\right| + \sup_{u_k\leq u\leq u_k+\varepsilon, u_k\in\mathcal{U}_k} \left|\langle M_2^o, X_j\rangle\right| \\ &\leq \sup_{u_k-\varepsilon\leq u\leq u_k, u_k\in\mathcal{U}_k} \left|\langle M_2^o, |X_j|\rangle\right| + \sup_{u_k\leq u\leq u_k+\varepsilon, u_k\in\mathcal{U}_k} \left|\langle M_2^o, |X_j|\rangle\right| \\ &\leq \max_{u_k\in\mathcal{U}_k} \left|\langle M_2^{(1)}, |X_j|\rangle\right| + \max_{u_k\in\mathcal{U}_k} \left|\langle M_2^{(2)}, |X_j|\rangle\right| \end{aligned}$$

The first inequality holds because in each of the two cases, all the elements in M_2^o have the same sign. So the inner product is maximized if the elements in X_j also have the same sign. The second inequality then follows because now the magnitude of the inner product is increasing in the magnitude of any elements in M_2^o.

I only consider $\max_{u_k\in\mathcal{U}_k}\left|\langle M_2^{(1)}, |X_j|\rangle\right|$ as the argument for $\max_{u_k\in\mathcal{U}_k}\left|\langle M_2^{(2)}, |X_j|\rangle\right|$ is identical. The expectation of a generic element in $M_2^{(1)}$ satisfies $\mu_k \equiv \mathbb{E}(\xi_{it}^{(1)}(u_k)) = \mathbb{P}(u_k - \varepsilon < U_{it} \leq u_k) = \varepsilon$. Let $\bar{M}_2^{(1)} = (\mu_k)_{1\leq i\leq N, 1\leq t\leq T}$. Then we have

$$\begin{aligned} \max_{u_k\in\mathcal{U}_k} \left|\langle M_2^{(1)}, |X_j|\rangle\right| &\leq \max_{u_k\in\mathcal{U}_k} \left|\langle M_2^{(1)} - \bar{M}_2^{(1)}, |X_j|\rangle\right| + \max_{u_k\in\mathcal{U}_k} \left|\langle \bar{M}_2^{(1)}, |X_j|\rangle\right| \\ &\leq \max_{u_k\in\mathcal{U}_k} \left|\langle M_2^{(1)} - \bar{M}_2^{(1)}, |X_j|\rangle\right| + \varepsilon\sqrt{NT}||X_j||_F \\ &\leq \max_{u_k\in\mathcal{U}_k} \left|\langle M_2^{(1)} - \bar{M}_2^{(1)}, |X_j|\rangle\right| + C\sqrt{NT} \end{aligned}$$

Finally, the first term is also bounded by $C_1\sqrt{NT\log(NT)}$ following exactly the same argument as for $\max_{u_k\in\mathcal{U}_K}\left|\left\langle \nabla\boldsymbol{\rho}_{u_k}(V(u_k)), X_j\right\rangle\right|$ because elements in $(M_2^{(1)} - \bar{M}_2^{(1)})$ are i.i.d, mean zero, and bounded in magnitude by 1. Therefore, as the bound does not depend on j and p

is a constant, equation (A.1) follows.

Proof of Equation (A.2). The proof follows similar argument as for equation (A.1). Again, let $\mathcal{U}_K = (u_1, u_2, ..., u_K)$ be an ε-net in $\mathcal{U}$. Let $\varepsilon = \frac{1}{\sqrt{N \vee T}}$. Then

$$\sup_{u \in \mathcal{U}} ||\nabla \boldsymbol{\rho}_u(V(u))|| \leq \max_{u_k \in \mathcal{U}_K} ||\nabla \boldsymbol{\rho}_{u_k}(V(u_k))|| + \sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} ||\nabla \boldsymbol{\rho}_u(V(u)) - \nabla \boldsymbol{\rho}_{u_k}(V(u_k))||$$

For the first term,

$$\begin{aligned}
\mathbb{P}\Big(\max_{u_k \in \mathcal{U}_K} ||\nabla \boldsymbol{\rho}_{u_k}(V(u_k))|| > C_2\sqrt{N \vee T}\Big) &\leq \max_{u_k \in \mathcal{U}_K} \mathbb{P}\Big(||\nabla \boldsymbol{\rho}_{u_k}(V(u_k))|| > C_2\sqrt{N \vee T}\Big) \\
&\leq \frac{C_2'}{\varepsilon} \exp(-C_2''\sqrt{NT}) \\
&= \frac{C_2''\sqrt{N \vee T}}{\exp(C_2''\sqrt{NT})}
\end{aligned}$$

where the second line is from Corollary 2.3.5 in Tao (2012) (p.129) that bounds the operator norm of a matrix with i.i.d., mean zero entries that are bounded in magnitude by 1.

For the second term, from the proof of equation (A.1),

$$\sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} ||\nabla \boldsymbol{\rho}_u(V(u)) - \nabla \boldsymbol{\rho}_{u_k}(V(u_k))|| \leq \sup_{|u-u_k| \leq \varepsilon, u_k \in \mathcal{U}_K} \Big(||M_1|| + ||M_2^o||\Big)$$

where M_1 and M_2^o are defined in the proof of equation (A.1). By definition, the operator norm of a generic matrix A is $\sup_{||x||_F=1} ||Ax||_F$ where x is a vector of unit Euclidean norm. When all the elements in A have the same sign, the supremum is achieved if all elements in x also have the same sign. Therefore, $\sup_{||x||_F=1} ||Ax||_F \leq \sup_{||x||_F=1} ||A \cdot |x|||_F$. Then for a matrix B such all elements in B also have the same sign and have weakly larger magnitude than those in A, $\sup_{||x||_F=1} ||A \cdot |x|||_F \leq \sup_{||x||_F=1} ||B \cdot |x|||_F = \sup_{||x||_F=1} ||Bx||_F = ||B||$. Hence,

$$\begin{aligned}
\sup_{\varepsilon, u_k \in \mathcal{U}_K} \Big(||M_1|| + ||M_2^o||\Big) \leq& \sup_{u_k - \varepsilon \leq u \leq u_k, u_k \in \mathcal{U}_k} \Big(||M_1|| + ||M_2^o||\Big) + \sup_{u_k \leq u \leq u_k + \varepsilon, u_k \in \mathcal{U}_k} \Big(||M_1|| + ||M_2^o||\Big) \\
\leq& \varepsilon||\mathbf{1}_{N \times T}|| + \sup_{u_k - \varepsilon \leq u \leq u_k, u_k \in \mathcal{U}_k} ||M_2^{(1)}|| + \sup_{u_k \leq u \leq u_k + \varepsilon, u_k \in \mathcal{U}_k} ||M_2^{(2)}|| \\
=& \varepsilon||\mathbf{1}_{N \times T}|| + \max_{u_k \in \mathcal{U}_k} ||M_2^{(1)}|| + \max_{u_k \in \mathcal{U}_k} ||M_2^{(2)}||
\end{aligned}$$

where $\mathbf{1}_{N \times T}$ is a constant matrix of all ones. It comes from M_1. $M_2^{(1)}$ and $M_2^{(2)}$ follow the same definition in the proof of equation (A.1).

Again, let us only consider $\max_{u_k \in \mathcal{U}_k} ||M_2^{(1)}||$.

$$\max_{u_k \in \mathcal{U}_k} ||M_2^{(1)}|| \leq \max_{u_k \in \mathcal{U}_k} ||M_2^{(1)} - \bar{M}_2^{(1)}|| + \max_{u_k \in \mathcal{U}_k} ||\bar{M}_2^{(1)}||$$

Note that elements in $M_2^{(1)} - \bar{M}_2^{(1)}$ are again i.i.d., mean zero, and bounded in magnitude by 1, so it has the same upper bound as $\max_{u_k \in \mathcal{U}_K} ||\nabla \boldsymbol{\rho}_{u_k}(V(u_k))||$. For the second term, by the same argument as in the proof of equation (A.1), $\bar{M}_2^1 = \varepsilon \mathbf{1}_{N\times T}$ where $\mathbf{1}_{N\times T}$ is an $N \times T$ matrix of all ones whose operator norm is $O(N \vee T)$. Therefore $\max_{u_k \in \mathcal{U}_k} ||\bar{M}_2^{(1)}|| \leq C_2''' \sqrt{N \vee T}$ by the choice of ε. This completes the proof. □

Proof of Lemma A2. For any $w_1 \in \mathbb{R}$, $\mathbb{1}(w_1 \leq z)$ is weakly increasing in z. Therefore, if $w_2 > 0$, $z \geq 0$, so $\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0) \geq 0$, the second inequality thus holds. Similarly,

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

Since $\kappa w_2 < w_1$, the right hand side is nonnegative. Hence the first inequality holds.

When $w_2 \leq 0$, note that $\int_0^{w_2} \Big(\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0)\Big) dz = \int_{w_2}^0 \Big(\mathbb{1}(w_1 \leq 0) - \mathbb{1}(w_1 \leq z)\Big) dz$ and $\int_0^{\kappa w_2} \Big(\mathbb{1}(w_1 \leq z) - \mathbb{1}(w_1 \leq 0)\Big) dz = \int_{\kappa w_2}^0 \Big(\mathbb{1}(w_1 \leq 0) - \mathbb{1}(w_1 \leq z)\Big) dz$. Now that $z \leq 0$, $\mathbb{1}(w_1 \leq 0) - \mathbb{1}(w_1 \leq z) \geq 0$. Therefore, following the same argument in the previous case, we obtain the desired result. □

References

Abrevaya, Jason, and Christian M. Dahl. 2008. "The effects of birth inputs on birthweight: evidence from quantile estimation on panel data." *Journal of Business & Economic Statistics*, 26(4): 379–397.

Agarwal, Alekh, Sahand Negahban, and Martin J. Wainwright. 2012. "Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions." *The Annals of Statistics*, 40(2): 1171–1197.

Ando, Tomohiro, and Jushan Bai. 2019. "Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity." *Journal of the American Statistical Association*. DOI: 10.1080/01621459.2018.1543598.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2018. "Matrix completion methods for causal panel data models." *arXiv preprint arXiv:1710.10251.*

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica*, 77(4): 1229–1279.

Bai, Jushan, and Junlong Feng. 2019. "Robust principal component analysis with non-sparse errors." *arXiv preprint arXiv:1902.08735.*

Bai, Jushan, and Serena Ng. 2019*a*. "Matrix completion, counterfactuals, and factor analysis of missing data." *arXiv preprint arXiv:1910.06677.*

Bai, Jushan, and Serena Ng. 2019*b*. "Rank regularized estimation of approximate factor models." *Journal of Econometrics*, 212(1): 78–96.

Belloni, Alexandre, and Victor Chernozhukov. 2011. "$\ell 1$-penalized quantile regression in high-dimensional sparse models." *The Annals of Statistics*, 39(1): 82–130.

Canay, Ivan A. 2011. "A simple approach to quantile regression for panel data." *The Econometrics Journal*, 14(3): 368–386.

Candès, Emmanuel J., and Benjamin Recht. 2009. "Exact matrix completion via convex optimization." *Foundations of Computational mathematics*, 9(6): 717.

Candès, Emmanuel J., Xiaodong Li, Yi Ma, and John Wright. 2011. "Robust principal component analysis?" *Journal of the ACM (JACM)*, 58(3): 11.

Chao, Shih-Kang, Wolfgang K. Härdle, and Ming Yuan. 2019. "Factorisable multi-task quantile regression." *arXiv preprint arXiv:1507.03833.*

Chen, Liang. 2019. "Two-step estimation of quantile panel data models with interactive fixed effects." *Working Paper.*

Chen, Liang, Juan J. Dolado, and Jesus Gonzalo. 2019. "Quantile factor models." *arXiv preprint arXiv:1911.02173.*

Chernozhukov, Victor, Christian Hansen, Yuan Liao, and Yinchu Zhu. 2019. "Inference for heterogeneous effects using low-rank estimations." *arXiv preprint arXiv:1812.08089.*

Galvao, Antonio F., and Kengo Kato. 2016. "Smoothed quantile regression for panel data." *Journal of Econometrics*, 193(1): 92–112.

Galvao, Antonio F., Carlos Lamarche, and Luiz Renato Lima. 2013. "Estimation of censored quantile regression for panel data with fixed effects." *Journal of the American Statistical Association*, 108(503): 1075–1089.

Ganesh, Arvind, John Wright, Xiaodong Li, Emmanuel J. Candes, and Yi Ma. 2010. "Dense error correction for low-rank matrices via principal component pursuit." In *2010 IEEE International Symposium on Information Theory.* 1513–1517.

Harding, Matthew, and Carlos Lamarche. 2014. "Estimating and testing a quantile regression model with interactive effects." *Journal of Econometrics*, 178: 101–113.

Hsu, Daniel, Sham M. Kakade, and Tong Zhang. 2011. "Robust matrix decomposition with sparse corruptions." *IEEE Transactions on Information Theory*, 57(11): 7221–7234.

Kato, Kengo, Antonio F. Galvao Jr, and Gabriel V. Montes-Rojas. 2012. "Asymptotics for panel quantile regression models with individual effects." *Journal of Econometrics*, 170(1): 76–91.

Knight, Keith. 1998. "Limiting distributions for L_1 regression estimators under general conditions." *Annals of Statistics*, 26(2): 755–770.

Koenker, Roger. 2004. "Quantile regression for longitudinal data." *Journal of Multivariate Analysis*, 91(1): 74–89.

Lamarche, Carlos. 2010. "Robust penalized quantile regression estimation for panel data." *Journal of Econometrics*, 157(2): 396–408.

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

Lin, Zhouchen, Minming Chen, and Yi Ma. 2010. "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices." *arXiv preprint arXiv:1009.5055.*

Moon, Hyungsik Roger, and Martin Weidner. 2015. "Linear regression for panel with unknown number of factors as interactive fixed effects." *Econometrica*, 83(4): 1543–1579.

Moon, Hyungsik Roger, and Martin Weidner. 2019. "Nuclear norm regularized estimation of panel regression models." *arXiv preprint arXiv:1810.10987.*

Negahban, Sahand, and Martin J. Wainwright. 2011. "Estimation of (near) low-rank matrices with noise and high-dimensional scaling." *The Annals of Statistics*, 39(2): 1069–1097.

Negahban, Sahand, and Martin J. Wainwright. 2012. "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise." *Journal of Machine Learning Research*, 13(May): 1665–1697.

Negahban, Sahand N., Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu. 2012. "A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers." *Statistical Science*, 27(4): 538–557.

Pesaran, M. Hashem. 2006. "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, 74(4): 967–1012.

Tao, Terence. 2012. *Topics in Random Matrix Theory.* Vol. 132 of *Graduate Studies in Mathematics*, American Mathematical Society.

van der Vaart, Aad W., and Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes.* Springer.

Yuan, Xiaoming, and Junfeng Yang. 2013. "Sparse and low-rank matrix decomposition via alternating direction method." *Pacific Journal of Optimization*, 9(1): 167.

Zhou, Zihan, Xiaodong Li, John Wright, Emmanuel Candes, and Yi Ma. 2010. "Stable principal component pursuit." In *2010 IEEE International Symposium on Information Theory.* 1518–1522.