WILD BOOTSTRAP INFERENCE FOR PENALIZED QUANTILE REGRESSION FOR LONGITUDINAL DATA*

CARLOS LAMARCHE AND THOMAS PARKER†

Abstract: The existing theory of penalized quantile regression for longitudinal data has focused primarily on point estimation. In this work, we investigate statistical inference. We propose a wild residual bootstrap procedure and show that it is asymptotically valid for approximating the distribution of the penalized estimator. The model puts no restrictions on individual effects, and the estimator achieves consistency by letting the shrinkage decay in importance asymptotically. The new method is easy to implement and simulation studies show that it has accurate small sample behavior in comparison with existing procedures. Finally, we illustrate the new approach using U.S. Census data to estimate a model that includes more than eighty thousand parameters.

Keywords: Quantile regression; panel data; penalized estimator; bootstrap inference.
JEL classification: C15; C21; C23.

1. Introduction

We consider a longitudinal data model of conditional quantiles with individual intercepts. Variations of this model have been extensively studied in the literature since at least Neyman and Scott (1948). Recent contributions to the literature using this model for quantile regression have emphasized the drawbacks of estimating a large number of individual intercepts (N) when the number of time periods (T) is small (see Galvao and Kato, 2018, for an excellent survey). Koenker (2004) proposed an estimator where N individual parameters are regularized by a Lasso-type penalty, shrinking them towards a common value. As in the case of the Gaussian random effect estimator, shrinkage can reduce the variability of the estimator of the slope parameter in the quantile regression model (Koenker, 2004). In models with short T, shrinkage can reduce the bias of the fixed effects estimator of the slope parameter as well (Harding and Lamarche, 2019).

Although the regularization procedure has advantages, the asymptotic distribution of the estimator is difficult to approximate. It is known that Lasso-type estimators have non-standard limiting distributions (Knight and Fu, 2000), but in the case of quantile regression, there are new challenges. Because individual intercepts are treated as parameters, the increasing dimension of the parameter vector as the number of units increases can be an issue. In the case of estimators without regularization, Kato, Galvao, and Montes-Rojas (2012) and Galvao, Gu, and Volgushev (2020) found that T must grow faster than N for consistency and asymptotic normality at rates that are, at best, similar to standard non-linear panel data models (Hahn and Newey, 2004). Second, the covariance matrix of quantile regression estimators typically depends on conditional densities and the penalized estimator of Koenker (2004)

*This draft: November 1, 2022.
†Corresponding author: Carlos Lamarche, Department of Economics, University of Kentucky, 223G Gatton College of Business & Economics, Lexington, KY 40506. Email: clamarche@uky.edu.
Thomas Parker: Department of Economics, University of Waterloo, 200 University Ave. West, Waterloo, ON, Canada N2L 3G1. Email: tmparker@uwaterloo.ca

is no exception. Inference based on the asymptotic distribution requires non-parametric estimation of nuisance parameters, which can lead to important size distortions (He, 2018).

Motivated by these limitations, cross-sectional pairs (or block) bootstrap, which samples sets of covariate and response vectors over individuals with replacement, appears to be a natural alternative method for inference. However, we demonstrate that the cross-sectional pairs bootstrap does not approximate well the limiting distribution of the penalized estimator. We consider instead a wild residual bootstrap procedure, which was previously employed by Feng, He, and Hu (2011), and Wang, Van Keilegom, and Maidman (2018) in cross-sectional settings. We investigate the application of the procedure to longitudinal data and show that the proposed wild bootstrap procedure is a consistent estimator of the distribution of the penalized estimator.

We begin by deriving consistency and asymptotic normality results for ℓ_1 penalized estimators of a longitudinal model in which individual effects can be correlated with the regressors. Although our model might be considered to be high-dimensional, the number of parameters is smaller than the number of observations, as in the pioneering work by Koenker (2004), and thus our results are obtained without assuming sparsity in terms of the individual intercepts. Consistency and asymptotic normality with T growing faster than N are achieved by letting the penalty parameter that controls shrinkage diminish in importance asymptotically. Thus, relative to Koenker (2004), the asymptotic bias of the estimator is zero in our case. The consistency and asymptotic normality results are new — they extend the heuristic results in Koenker (2004) obtained for a model with individual effects as location shifts and are not included in Kato, Galvao, and Montes-Rojas (2012) and Galvao, Gu, and Volgushev (2020) because they did not consider penalized estimation.

The main theoretical contribution is to show that the distribution of the wild bootstrap estimator consistently estimates the asymptotic distribution and covariance of the penalized estimator. The results include the special case of no penalization, and thus, these results also show the consistency of the wild bootstrap for the quantile regression estimator with fixed effects. The consistency of the wild bootstrap is established using developments that are critically different to those used in Wang, Van Keilegom, and Maidman (2018). We also consider bootstrap estimation of the asymptotic covariance matrix of the slope parameter estimator, which is novel in the panel quantile literature. As emphasized in Gonçalves and White (2005), Hagemann (2017), and Hahn and Liao (2021), the weak convergence of the bootstrap estimator does not necessarily imply convergence of the bootstrap second moment estimator. Therefore, we provide conditions and establish a result that supports using the second moment of the bootstrap distribution to estimate the asymptotic variance of the estimator.

Several penalized estimators for quantile regression models have been proposed in the literature since Koenker (2004). Belloni and Chernozhukov (2011) propose quantile regression estimators for high-dimensional sparse models using cross-sectional data. Wang (2013) considers a penalized least absolute deviation estimator, and Wang (2019) derives error bounds for the penalized estimator under weak conditions. Lamarche (2010) investigates the selection of a regularization parameter, and Lee, Liao, Seo, and Shin (2018) study estimation of a high-dimensional quantile regression model with a change point, or threshold. Harding and Lamarche (2017, 2019) investigate estimation of models with attrition and correlated random effects. Gu and Volgushev (2019) propose a method for estimation of models with unknown group membership. Chen and Pouzo (2009) establish the validity of a related weighted bootstrap procedure for the limiting distribution of a penalized sieve estimator and consider applications using quantile regression (see also Chen and Pouzo, 2015). The literature on penalized estimation methods for linear panel data models has also grown in the last decade (see, e.g., Kock, 2013, 2016; Belloni, Chernozhukov, Hansen, and Kozbur, 2016; Su, Shi, and Phillips, 2016; Su and Ju, 2018; Caner and Kock, 2018; Kock and Tang, 2019, among others.)

This paper is organized as follows. The next section provides background and discusses the motivation of our study. It also introduces the proposed wild residual bootstrap approach. Section 3 presents theoretical results. Section 4 investigates the small sample performance of the method, showing that the estimator has satisfactory performance under different specifications and it performs better than the cross-sectional pairs bootstrap procedure. Section 5 presents extensions to the basic model. Section 6 illustrates the theory and provides practical guidelines from an application of the method. Considering data from the U.S. Census, we estimate a quantile function with more than eighty thousand parameters to study how wages of U.S. workers have been affected by the North American Free Trade Agreement. Finally, Section 7 concludes. One appendix contains proof of the main results, while a supplementary appendix contains additional technical results and proofs.

2. Inference for penalized quantile regression

2.1. **Background and Motivation.** We observe repeated measures $\{(y_{it}, \boldsymbol{x}_{it}')\}_{t=1}^{T}$ for each subject $1 \leq i \leq N$. The variable $y_{it} \in \mathbb{R}$ denotes the response for i at time t and $\boldsymbol{x}_{it}$ denotes a p-dimensional vector of covariates. Although the number of repeated observations does not vary with i, the analysis can be trivially extended to consider T_i as long as $\max T_i / \min T_i$ is bounded for $1 \leq i \leq N$ (Gu and Volgushev, 2019). The model considered in this paper is

$$Q_y(\tau|\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}'\boldsymbol{\beta}_0(\tau) + \alpha_{i0}(\tau), \tag{2.1}$$

where $\tau \in (0,1)$ and $Q_y(\tau|\boldsymbol{x}_{it})$ is the τ-th quantile of the conditional distribution of y_{it} given $\boldsymbol{x}_{it}$. It is assumed that the vector $\boldsymbol{x}_{it}$ does not contain an intercept. The parameter of interest is $\boldsymbol{\beta}_0(\tau) \in \mathbb{R}^p$ and $\alpha_{i0}(\tau)$ is treated as a nuisance parameter. Because we consider just one value of τ, we suppress the dependence of the parameters on τ in the sequel.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')' \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+N}$, where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$, and let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0')'$. To estimate $\boldsymbol{\theta}_0$, we consider the following estimator:

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}')' = \operatorname*{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i) + \lambda_T \sum_{i=1}^{N} |\alpha_i|, \tag{2.2}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile regression loss function. The tuning parameter $\lambda_T \geq 0$ depends on T and it can also depend on data, as discussed below.

The penalty term in (2.2) helps improve the finite sample performance of the fixed effects estimator, which is defined for $\lambda_T = 0$. Shrinkage of the individual effects can lead to reductions of the variance of the estimator. In models with incidental parameters, the penalty term reduces the noise in the estimation of individual intercepts, and consequently, it can also reduce the bias of the fixed effects estimator of $\boldsymbol{\beta}_0$. The online appendix presents simulation evidence to illustrate finite sample improvements when the time dimension is short, complementing the evidence presented in Koenker (2004) and Harding and Lamarche (2019). See Bester and Hansen (2009) for a related penalty approach to bias reduction in nonlinear models with fixed effects.

We establish conditions that result in a tractable asymptotic distribution for the estimator defined in (2.2). However, we expect that resampling methods offer a more accurate description of the distribution of the estimator in finite samples. In practice, the cross-sectional pairs bootstrap, which samples over i with replacement keeping the entire block of time series observations for each i, has been used as a method for inference, primarily in the fixed effects case when $\lambda_T = 0$. However, the cross-sectional pairs bootstrap does not provide a good approximation to the sampling distribution of the penalized estimator (2.2), as in the case of the pairs bootstrap procedure for the Lasso estimator (Camponovo, 2015).

2.2. **A cross-sectional pairs bootstrap procedure.** We now offer a heuristic illustration of some problems with using a cross-sectional pairs bootstrap and the penalized quantile regression estimator. The cross-sectional pairs bootstrap can be used successfully to estimate the distribution of the quantile regression model with unpenalized fixed effects, but it will be shown below that the penalty causes problems for this approach to resampling. We fix N in this section to avoid the effect of a diverging number of parameters as the sample size increases (later, asymptotic approximations will be found assuming that T grows faster than N). This allows us to see problems with the cross-sectional pairs without the additional incidental parameters problem. Define $\boldsymbol{\gamma} = (\boldsymbol{\delta}', \boldsymbol{\eta}')' \in \mathbb{R}^{p+N}$, where $\boldsymbol{\delta} = \sqrt{NT}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and for $i = 1, \ldots N$, $\eta_i = \sqrt{T}(\alpha_i - \alpha_{i0})$. Then let

$$\mathbb{V}_T(\boldsymbol{\gamma}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_\tau \left(u_{it} - \frac{\boldsymbol{x}_{it}'\boldsymbol{\delta}}{\sqrt{NT}} - \frac{\eta_i}{\sqrt{T}} \right) - \rho_\tau(u_{it}) \right\} + \lambda_T \sum_{i=1}^{N} \left\{ \left| \alpha_{i0} + \frac{\eta_i}{\sqrt{T}} \right| - |\alpha_{i0}| \right\}, \tag{2.3}$$

where $u_{it} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 - \alpha_{i0}$. This objective function is equivalent to (2.2). Knight and Fu (2000) developed a method for dealing with the asymptotic behavior of this objective function, stated here as a lemma.

Lemma 1 (Knight and Fu (2000))**.** *Under Assumptions B1-B5 below, if N is fixed, $T \to \infty$ and $\lambda_T/\sqrt{T} \to \lambda_0 \geq 0$, the minimizer of* (2.3)*, $\hat{\boldsymbol{\gamma}}$, converges weakly to the minimizer of $\mathbb{V} : \mathbb{R}^{p+N} \to \mathbb{R}$ defined by*

$$\mathbb{V}(\boldsymbol{\gamma}) = -\boldsymbol{\gamma}'\boldsymbol{B} + \frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{D}_1\boldsymbol{\gamma} + \lambda_0 \sum_{i=1}^{N} \left(\eta_i \operatorname{sgn}(\alpha_{i0}) I(\alpha_{i0} \neq 0) + |\eta_i| I(\alpha_{i0} = 0) \right),$$

where $\boldsymbol{D}_1$ is positive definite and $\boldsymbol{B} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}_0)$.

To examine the validity of the cross-sectional pairs bootstrap, consider an analog loss function for resampled data. Letting $\boldsymbol{y}_i$ and $\boldsymbol{X}_i$ denote the vector and matrix of response and covariate observations corresponding to unit i, a cross-sectional pairs bootstrap procedure resamples N pairs $(\boldsymbol{y}_i, \boldsymbol{X}_i)$ for

$1 \leq i \leq N$ with replacement. Let n_i^* denote the number of times unit i is redrawn from the original sample. Thus, the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ is approximated with $\tilde{\boldsymbol{\gamma}} = (\sqrt{NT}(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})', \sqrt{T}(\tilde{\boldsymbol{\alpha}}-\hat{\boldsymbol{\alpha}})')'$ where

$$\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\alpha}}'\right)' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^{N} n_i^* \sum_{t=1}^{T} \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i\right) + \lambda_T \sum_{i=1}^{N} n_i^* |\alpha_i|. \tag{2.4}$$

Since n_i^* is a multinomial weight with probability $1/N$, it is straightforward to calculate that the expected value of the objective function with respect to the bootstrap weights (i.e., conditional on the observations) is minimized at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$. However, a finite sample problem is associated with the presence of the penalty in the objective function. To see this, let $\alpha_i^* = n_i^*|\alpha_i|$ and $\mathcal{A} = \{i : \alpha_i^* \neq 0\}$ denote the "active" set corresponding to the penalty term in (2.4). In each bootstrap repetition, the cardinality of $\mathcal{A} < N$, leading to solutions $\tilde{\boldsymbol{\theta}}$ that can be potentially very different than the minimizer $\hat{\boldsymbol{\theta}}$. This may be especially so when α_i is correlated with $\boldsymbol{x}_{it}$.

To see other problems with the cross-sectional bootstrap, we can find the weak limit of the bootstrap objective function (2.5) similarly to Lemma 1. When we recenter (2.4) employing $\hat{\boldsymbol{\theta}}$, using the i chosen by resampling, we find a naive bootstrap analog of the original objective function (2.3), denoting $\hat{u}_{it} = y_{it} - \hat{\boldsymbol{\beta}}'\boldsymbol{x}_{it} - \hat{\alpha}_i$:

$$\tilde{\mathbb{V}}_T(\boldsymbol{\gamma}) = \sum_{i=1}^{N} n_i^* \sum_{t=1}^{T} \left\{\rho_\tau\left(\hat{u}_{it} - \frac{\boldsymbol{\delta}'\boldsymbol{x}_{it}}{\sqrt{NT}} - \frac{\eta_i}{\sqrt{T}}\right) - \rho_\tau(\hat{u}_{it})\right\} + \lambda_T \sum_{i=1}^{N} n_i^* \left\{\left|\hat{\alpha}_i + \frac{\eta_i}{\sqrt{T}}\right| - |\hat{\alpha}_i|\right\}. \tag{2.5}$$

As $T \to \infty$, assuming $\hat{\eta}_i = \sqrt{T}(\hat{\alpha}_i - \alpha_{i0}) \xrightarrow{d} A_i$ for $i = 1, \ldots N$ as $T \to \infty$, $\tilde{\mathbb{V}}_T$ converges weakly to

$$\tilde{\mathbb{V}}(\boldsymbol{\gamma}) = -\boldsymbol{\gamma}'\tilde{\boldsymbol{B}} + \frac{1}{2}\boldsymbol{\gamma}'\tilde{\boldsymbol{D}}_1\boldsymbol{\gamma} + \lambda_0 \sum_{i=1}^{N} n_i^*\big(\eta_i \operatorname{sgn}(\alpha_{i0}) I(\alpha_{i0} \neq 0) + (|\eta_i + A_i| - |A_i|)\, I(\alpha_{i0} = 0)\big).$$

However, there are two key differences with the resulting expression. The first problem with this limiting objective function is that $\tilde{\boldsymbol{B}} \neq \boldsymbol{B}$ and $\tilde{\boldsymbol{D}}_1 \neq \boldsymbol{D}_1$ from Lemma 1, due to the fact that recentering uses $\hat{\boldsymbol{\theta}}$, which is asymptotically biased if $\lambda_0 > 0$. Second, there is additional randomness arising from variable selection and resampling. (In the online appendix, we illustrate these issues with fixed N and T). In the next section, we propose a wild residual bootstrap that does not suffer from these shortcomings. Then we expect that the distribution of the wild bootstrap estimator $\boldsymbol{\gamma}^*$ provides a better approximation to the distribution of $\hat{\boldsymbol{\gamma}}$ in Lemma 1.

2.3. **Wild bootstrap procedures.** Let $\hat{u}_{it} = y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}} - \hat{\alpha}_i$ be the τ-th quantile residual. Let $u_{it}^* = w_{it}|\hat{u}_{it}|$ denote bootstrap residuals, where w_{it} is drawn randomly from a pre-determined distribution G_W that satisfies the following conditions:

A1. The τ-th quantile of G_W is equal to zero, i.e. $G_W(0) = \tau$.

A2. The support of G_W is bounded and contained in the interval $(-\infty, -c_1] \cup [c_2, \infty)$, where $c_1 > 0$ and $c_2 > 0$.

A3. The weight distribution G_W satisfies $-\int_{-\infty}^{0} w^{-1} dG_W(w) = \int_{0}^{+\infty} w^{-1} dG_W(w) = \frac{1}{2}$.

Several weight distributions have been proposed in the quantile regression literature that satisfy these conditions. Feng, He, and Hu (2011) propose, for $1/8 \le \tau \le 7/8$, the continuous weight density $g_W(w) = -wI(-2\tau - 1/4 \le w \le -2\tau + 1/4) + wI(2(1-\tau) - 1/4 \le w \le 2(1-\tau) + 1/4)$. Another distribution that satisfies A1-A3 is the two-point distribution at $w = 2(1-\tau)$ with probability τ and at $w = -2\tau$ with probability $(1-\tau)$. We adopt this distribution in the numerical examples. See Appendix 3 in Wang, Van Keilegom, and Maidman (2018) for additional examples of the weight distribution.

Using the bootstrap sample of residuals and the penalized quantile estimator as defined in equation (2.2), we can form $y^*_{it} = \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}} + \hat{\alpha}_i + u^*_{it}$ to obtain the bootstrap estimator:

$$\boldsymbol{\theta}^* = (\boldsymbol{\beta}^{*'}, \boldsymbol{\alpha}^{*'})' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(y^*_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta} - \alpha_i) + \lambda_T \sum_{i=1}^{N} |\alpha_i|. \tag{2.6}$$

Given a bootstrap sample $\{\boldsymbol{\beta}^*_b\}_{b=1}^{B}$, we can obtain confidence intervals that are asymptotically valid, as demonstrated in Theorem 3 below. Let $G^*_j(\alpha/2)$ and $G^*_j(1-\alpha/2)$ be the $(\alpha/2)$-th quantile and $(1-\alpha/2)$-th quantile of the bootstrap distribution of $\sqrt{NT}(\beta^*_j - \hat{\beta}_j)$ for $j = 1, 2, \ldots, p$. We obtain asymptotically valid $100(1-\alpha)\%$ confidence intervals for β_j by $[\hat{\beta}_j - (NT)^{-1/2}G^*_j(1-\alpha/2), \hat{\beta}_j - (NT)^{-1/2}G^*_j(\alpha/2)]$. Alternatively, Theorem 4 shows that we may also estimate the covariance matrix of $\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ using the estimated covariance matrix from the bootstrap sample, which can be used to estimate the variance without requiring density estimation and to construct bootstrap-t statistics for inference.

We may also consider a threshold estimator for $1 \le i \le N$, $\alpha^{**}_i = \hat{\alpha}_i I(|\hat{\alpha}_i| \ge a_T)$, where a_T is a constant that satisfies $a_T \to 0$ as $T \to \infty$. Define $v^*_{it} = w_{it}|\hat{v}_{it}|$, where $\hat{v}_{it} = y_{it} - \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}} - \alpha^{**}_i$. The response variable is generated as $y^{**}_{it} = \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}} + \alpha^{**}_i + v^*_{it}$, and the threshold estimator is defined as

$$\boldsymbol{\theta}^{**} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(y^{**}_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta} - \alpha_i) + \lambda_T \sum_{i=1}^{N} |\alpha_i|. \tag{2.7}$$

As in the case of the estimator defined in (2.6), we estimate the distribution of $\hat{\boldsymbol{\theta}}$ based on the estimator $\boldsymbol{\theta}^{**}$. Given the similarities between estimators (2.6) and (2.7), we derive below consistency and asymptotic normality results for (2.6) only. The performance of the bootstrap with this estimator is examined in the online appendix.

Remark 1. As in the case of other bootstrap estimators in quantile regression (Feng, He, and Hu, 2011; Hagemann, 2017), an adjustment based on the Bahadur representation of the estimator is recommended. For instance, under i.i.d. errors, residuals might be replaced by $\hat{u}_{it} + \sum_{i=1}^{N}\sum_{t=1}^{T} H_{it}(\tau - I(\hat{u}_{it} < 0))/\hat{f}(0)$, where $H_{it} = \boldsymbol{h}'_{it}(\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{h}_{it}\boldsymbol{h}'_{it})^{-1}\boldsymbol{h}_{it}$, $\boldsymbol{h}_{it} = (\boldsymbol{x}'_{it}, z_i)'$, z_i is an indicator variable for subject i, and $\hat{f}$ is an estimate of the density function defined below in B4. The exact form of the Bahadur representation for the panel quantile problem is given in Theorem 2 and details on the adjustment are given in Section 4.

2.4. **Tuning parameter selection.** The tuning parameter λ_T controls the degree of shrinkage of the individual effect α_i towards zero and the penalty helps to control the bias and variance of $\hat{\boldsymbol{\beta}}$. We restrict the tuning parameter to $\lambda_T \in \mathcal{L} \subset [0, \lambda_U]$, where λ_U is an upper bound. As shown in Lemma S.1 in the supplementary appendix, $\lambda_U = \max\{\tau, 1-\tau\}T$ is a natural choice because if λ_T is set larger than this value, all the individual effects will be set equal to zero. If the number of observed time periods T_i vary over i, then one would need to replace the T in these bounds with $\max_i T_i$. This estimator accommodates the choice of $\lambda_T = 0$, which means that the results below continue to hold for the corresponding unpenalized estimator.

The selection λ_T in related settings has been investigated in several papers (see, e.g., Lamarche, 2010; Lee, Noh, and Park, 2014; Wang, Van Keilegom, and Maidman, 2018). We follow Wang, Van Keilegom, and Maidman (2018) and employ cross-validation for tuning parameter selection. To the best of our knowledge, theory has not yet been developed for the stochastic order of λ_T when chosen using cross-validation, but in extensive simulations we have found that it tends to grow much more slowly than T, as required in Theorems 1 and 2 below.

3. Asymptotic theory

This section investigates the large sample properties of the proposed estimator. We consider the following assumptions:

B1. Suppose that $\{(y_{it}, \boldsymbol{x}_{it}) : t \geq 1\}$ are independent across i and independent and identically distributed (i.i.d.) within each unit i.

B2. For each $\phi > 0$,

$$\inf_{i \geq 1} \inf_{\|\boldsymbol{\theta}_i\|_1 = \phi} \mathrm{E}\left[\int_0^{(\alpha_i - \alpha_{i0}) + \boldsymbol{x}_{it}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)} (F_i(s|\boldsymbol{x}_{it}) - \tau)\,\mathrm{d}s\right] = \epsilon_\phi > 0,$$

where $F_i := F_{u_{it}|\boldsymbol{x}_{it}}$ is the distribution function of $u_{it} = y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0$ conditional on $\boldsymbol{x}_{it}$.

B3. The covariate vector $\boldsymbol{x}_{it}$ satisfies $\sup_{i,t} \|\boldsymbol{x}_{it}\| < M < \infty$ a.s.

These conditions are standard in the literature on quantile regression with individual effects. Conditions B1 and B2 are the same as Assumptions (A1) and (A3) in Kato, Galvao, and Montes-Rojas (2012). Condition B1 is relaxed in Kato et al. (2012) and in Section 5 below to allow for time dependence. Condition B2 is an identification condition and it is sufficient for consistency. Slightly weaker than the assumption that F_i has a continuous density given $\boldsymbol{x}_{it}$, it allows an expansion that guarantees the convexity of the limiting objective function, and therefore, the uniqueness of $(\boldsymbol{\beta}_0', \alpha_{i0})$ for all $1 \leq i \leq N$. Assumption B3 is a simple way to assume appropriate moment conditions on the covariates and it is similar to (B1) in Kato, Galvao, and Montes-Rojas (2012) and (A1) in Gu and Volgushev (2019). The condition can be relaxed as in Kato, Galvao and Montes-Rojas (2012). Condition B3 can be replaced with the moment condition $\sup_{i \geq 1} \mathrm{E}\left[\|\boldsymbol{x}_{i1}\|^{2s}\right] < \infty$ for some $s \geq 1$. The implication of this weaker condition is that $N/T^s \to 0$ instead of $\log(N)/T \to 0$ to achieve consistency, as demonstrated in Theorem 1.

The consistency of the estimator $\hat{\boldsymbol{\theta}}$ is needed to establish the main result stated in Theorem 3.

Theorem 1. *Under Assumptions B1-B3, if* $\log(N)/T \to 0$ *and* $\lambda_T = o_p(T)$ *as* $N, T \to \infty$*, then the estimator* $\hat{\boldsymbol{\theta}}$ *defined in equation* (2.2) *is a consistent estimator of* $\boldsymbol{\theta}_0$.

Remark 2. Theorem 1 is of independent interest as it has not been established the consistency of the penalized estimator under arbitrary dependence between regressors and individual effects. The result depends on the condition that λ_T, the parameter governing penalization of the individual effects, grows slowly as T increases.

We now focus our attention on weak convergence and we present a series of results to facilitate the estimation of standard errors and confidence intervals. To show asymptotic normality of the estimator, it is necessary to strengthen the conditions required for consistency slightly with the following conditions routinely adopted in the panel quantile regression literature (see, e.g., assumptions (B2) and (B3) in Kato, Galvao, and Montes-Rojas, 2012, and assumption (A2) in Gu and Volgushev, 2019).

B4. The conditional density function $f_i := f_{u_{it}|\boldsymbol{x}_{it}}$ corresponding to F_i is uniformly bounded and has a bounded first derivative $\overline{f} := \sup_i \sup_{u\in\mathbb{R},\boldsymbol{x}\in\mathbb{R}^p} |f_i(u|\boldsymbol{x})| < \infty$ and $\overline{f'} := \sup_i \sup_{u\in\mathbb{R},\boldsymbol{x}\in\mathbb{R}^p} |f_i'(u|\boldsymbol{x})| < \infty$. Assume that in an open neighborhood $\mathcal{U}$ of 0, f_i is bounded away from zero for all realizations of $\boldsymbol{x}_{it}$:

$$\underline{f} := \inf_i \inf_{u\in\mathcal{U},\boldsymbol{x}\in\mathbb{R}^p} |f_i(u|\boldsymbol{x})| < \infty.$$

B5. Let $\varphi_i := \mathrm{E}\left[f_i(0|\boldsymbol{x}_{i1})\right]$, $\boldsymbol{E}_i := \mathrm{E}\left[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\right]$ and $\boldsymbol{J}_i := \mathrm{E}\left[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}'\right]$. Let

$$\boldsymbol{D}_N = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{J}_i - \varphi_i^{-1}\boldsymbol{E}_i\boldsymbol{E}_i'\right).$$

Suppose that $\boldsymbol{D}_N$ is positive definite for all N and there is a positive definite matrix $\boldsymbol{D}$ such that $\boldsymbol{D} = \lim_{N\to\infty}\boldsymbol{D}_N$. Also assume that

$$\boldsymbol{V} = \tau(1-\tau)\times\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[\left(\boldsymbol{x}_{i1} - \varphi_i^{-1}\boldsymbol{E}_i\right)\left(\boldsymbol{x}_{i1} - \varphi_i^{-1}\boldsymbol{E}_i\right)'\right]$$

is positive definite.

Then we have the following result:

Theorem 2. *Under Assumptions B1-B5, if* $N^2(\log N)^3/T \to 0$ *and* $\lambda_T = o_p(T^{1/2}(\log N)^{1/2})$ *as* $N, T \to \infty$*, then*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = \boldsymbol{D}^{-1}\boldsymbol{V}\boldsymbol{D}^{-1}$.

Remark 3. As with Condition G in Theorem 3.2 in Gu and Volgushev (2019), Theorem 2 provides a selection rule for candidate values of the tuning parameters that are justified by theory. The limiting distribution for this estimator matches that of the conventional fixed effects estimator derived in Kato, Galvao, and Montes-Rojas (2012) because the tuning parameter λ_T diverges at a slow rate.

Remark 4. Because the goal of the shrinkage estimator here is not variable selection but regularization of the estimated $\hat{\alpha}_i$, the rate of growth of λ_T is different than what would usually be used in high-dimensional models (Belloni and Chernozhukov (2011, p. 86), Lee, Liao, Seo, and Shin (2018, eq.

2.3), and Wang (2019, Theorem 3.2)). This difference in stochastic order is because the individual effects $\{\alpha_{i0}\}_i$ are not assumed sparse and this condition on λ_T is needed for consistency in models with regressors correlated with individual latent effects. Moreover, perhaps not surprisingly, the rates derived for linear models (see, e.g., Kock, 2013, 2016) are also different to the rate required for establishing the asymptotic normality of the quantile estimator.

The wild residual bootstrap procedure is consistent as an estimator of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, as the next theorem shows.

Theorem 3. *Under Assumptions A1-A3 and the conditions of Theorem 2,*

$$\sup_{b\in\mathbb{R}^p}\left|\mathrm{P}\left\{\sqrt{NT}(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}})\leq b|\boldsymbol{S}\right\}-\mathrm{P}\left\{\sqrt{NT}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\leq b\right\}\right|\xrightarrow{p}0$$

where $\boldsymbol{S}$ *denotes the observed sample and* $\boldsymbol{\beta}^*$ *denotes the slope estimator defined by* (2.6).

Remark 5. By setting $\lambda_T = 0$, Theorem 3 also implies consistency of the wild residual bootstrap for the unpenalized estimator with individual effects and i.i.d. errors.

Remark 6. The results allow for a data-dependent λ_T but they do not allow selecting the tuning parameter at each bootstrap repetition. While theoretical developments are out of the scope of this paper, we investigated if this idea leads to improvements in the finite sample performance of the estimator. We did not find significant changes relative to the results presented in Section 4, although the computational cost of the procedure is higher.

Theorem 3 only shows consistency of the bootstrap distribution estimator. Theorem 4 ahead shows that the bootstrap covariance matrix, defined as

$$\boldsymbol{\Omega}^*=\mathrm{E}^*\left[NT\left(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}\right)\left(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}\right)'\right],$$

may be used to estimate the covariance of $\sqrt{NT}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$. In practice, one simply uses the sample covariance of all the bootstrap repetitions, increasing the number of repetitions to bring the sample average as close as desired to the bootstrap expectation. Variance estimation using the bootstrap was formally investigated for quantile regression with clustered data in Hagemann (2017), but the model in this paper is complicated by the diverging number of individual effects as $N\to\infty$ and the penalty term in (2.6).

Theorem 4. *Under Assumptions A1-A3 and the conditions of Theorem 2, if* $\boldsymbol{\theta}_i$ *for* $1\leq i\leq N$ *lie in a compact set and* $\sup_{N,T}\mathrm{E}\left[|\sqrt{N}\lambda_T/\sqrt{T}|^q\right]<\infty$ *for* $q>2$*, then* $\|\boldsymbol{\Omega}^*-\boldsymbol{\Omega}\|\xrightarrow{p}0$.

The assumptions that are required for Theorem 4 are slightly stronger than those used in Theorem 3. The requirement on λ_T is due to its presence in asymptotic expansions leading to the Bahadur representation of $\boldsymbol{\beta}^*$ and is similar to the moment requirement made on the covariates in Hagemann (2017). The compactness assumption must be made to ensure that expansions used in the asymptotic approximation are uniformly bounded.

		Quantile 0.5						Quantile 0.75					
		Method:			Method:			Method:			Method:		
N	T	CS	WB		CS	WB		CS	WB		CS	WB	
		PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE
		Location shift model ($\zeta = 0$) and $u \sim \mathcal{N}(0,1)$											
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.697	0.902	0.905	0.675	0.909	0.913	0.702	0.854	0.854	0.640	0.848	0.859
100	10	0.720	0.908	0.868	0.683	0.915	0.876	0.737	0.887	0.900	0.683	0.885	0.903
200	5	0.677	0.923	0.925	0.670	0.916	0.920	0.668	0.862	0.861	0.641	0.873	0.877
200	10	0.650	0.925	0.857	0.700	0.927	0.870	0.662	0.898	0.909	0.691	0.897	0.915
25	50	0.886	0.908	0.904	0.779	0.887	0.882	0.857	0.881	0.880	0.758	0.885	0.888
25	100	0.903	0.910	0.911	0.811	0.902	0.905	0.908	0.898	0.901	0.816	0.892	0.898
50	50	0.833	0.905	0.903	0.769	0.884	0.880	0.831	0.893	0.888	0.768	0.889	0.892
50	100	0.847	0.900	0.895	0.831	0.903	0.900	0.854	0.902	0.897	0.827	0.898	0.902
		Location shift model ($\zeta = 0$) and $u \sim t_3$											
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.710	0.906	0.911	0.674	0.902	0.912	0.701	0.828	0.833	0.623	0.819	0.840
100	10	0.713	0.923	0.880	0.660	0.919	0.881	0.734	0.881	0.893	0.680	0.864	0.880
200	5	0.681	0.932	0.936	0.645	0.907	0.927	0.669	0.841	0.852	0.573	0.816	0.852
200	10	0.650	0.922	0.834	0.661	0.931	0.845	0.641	0.881	0.892	0.678	0.859	0.889
25	50	0.887	0.921	0.916	0.784	0.906	0.906	0.852	0.887	0.886	0.738	0.881	0.881
25	100	0.905	0.901	0.901	0.819	0.892	0.891	0.870	0.883	0.891	0.801	0.891	0.900
50	50	0.850	0.898	0.895	0.759	0.884	0.884	0.839	0.886	0.889	0.761	0.900	0.895
50	100	0.848	0.886	0.887	0.816	0.899	0.892	0.837	0.885	0.890	0.770	0.863	0.875
		Location shift model ($\zeta = 0$) and $u \sim \chi^2_3$											
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.730	0.906	0.912	0.633	0.897	0.907	0.690	0.728	0.745	0.589	0.716	0.716
100	10	0.673	0.894	0.871	0.647	0.870	0.855	0.718	0.828	0.844	0.707	0.827	0.838
200	5	0.708	0.920	0.927	0.657	0.915	0.910	0.652	0.745	0.753	0.602	0.763	0.763
200	10	0.686	0.901	0.860	0.642	0.880	0.838	0.703	0.833	0.853	0.703	0.818	0.831
25	50	0.791	0.882	0.883	0.716	0.891	0.896	0.749	0.839	0.839	0.707	0.861	0.862
25	100	0.845	0.887	0.890	0.749	0.873	0.874	0.781	0.871	0.879	0.683	0.856	0.860
50	50	0.768	0.869	0.872	0.726	0.898	0.897	0.758	0.862	0.862	0.727	0.876	0.877
50	100	0.798	0.879	0.880	0.735	0.891	0.892	0.770	0.861	0.867	0.710	0.864	0.877

TABLE 4.1. *Empirical coverage probabilities of the bootstrap confidence interval for a nominal 90% level. CS denotes cross-sectional pairs bootstrap, WB denotes wild bootstrap, PQR denotes the penalized estimator, and FE is the unpenalized fixed effects estimator.*

4. Simulation Study

In this section, we report the results of several simulation experiments designed to evaluate the performance of the method in finite samples. We consider a data generating process similar to the ones considered in Koenker (2004) and Kato, Galvao and Montes-Rojas (2012). The dependent variable is $y_{it} = \alpha_i + x_{it} + (1 + \zeta x_{it})u_{it}$, where $x_{it} = 0.5\alpha_i + z_i + \epsilon_{it}$, and z_i and ϵ_{it} are i.i.d. random variables distributed as χ^2 with 3 degrees of freedom (χ^2_3). The corresponding quantile regression function is $Q_y(\tau|x_{it}) = \alpha_{0i} + \beta_0 x_{it}$, where $\alpha_{0i} = \alpha_i + F_u(\tau)^{-1}$, $\beta_0 = 1 + \zeta F_u(\tau)^{-1}$, and $F_u(\cdot)$ denotes the distribution of the error term, u_{it}.

		Quantile 0.5						Quantile 0.75					
		Method:			Method:			Method:			Method:		
N	T	CS	WB		CS	WB		CS	WB		CS	WB	
		PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE
Location-scale shift model ($\zeta = 0.5$) and $u \sim \mathcal{N}(0,1)$													
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.688	0.861	0.881	0.699	0.884	0.893	0.592	0.798	0.824	0.690	0.819	0.827
100	10	0.619	0.899	0.865	0.657	0.896	0.868	0.611	0.864	0.867	0.652	0.877	0.887
200	5	0.660	0.871	0.892	0.674	0.868	0.892	0.540	0.825	0.823	0.605	0.850	0.848
200	10	0.650	0.913	0.852	0.651	0.904	0.860	0.567	0.876	0.880	0.627	0.868	0.884
25	50	0.698	0.898	0.893	0.676	0.900	0.899	0.696	0.881	0.882	0.672	0.866	0.865
25	100	0.762	0.906	0.906	0.678	0.894	0.896	0.749	0.894	0.897	0.678	0.892	0.893
50	50	0.685	0.896	0.892	0.678	0.888	0.883	0.679	0.888	0.889	0.669	0.886	0.888
50	100	0.720	0.900	0.902	0.674	0.905	0.903	0.697	0.879	0.881	0.677	0.901	0.902
Location-scale shift model ($\zeta = 0.5$) and $u \sim t_3$													
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.723	0.857	0.886	0.758	0.866	0.890	0.620	0.751	0.769	0.721	0.769	0.781
100	10	0.664	0.902	0.882	0.676	0.895	0.873	0.638	0.862	0.874	0.672	0.855	0.857
200	5	0.754	0.875	0.910	0.722	0.873	0.909	0.513	0.767	0.775	0.707	0.780	0.795
200	10	0.650	0.899	0.857	0.649	0.910	0.843	0.554	0.816	0.835	0.634	0.822	0.824
25	50	0.742	0.910	0.910	0.682	0.914	0.916	0.694	0.874	0.875	0.672	0.869	0.873
25	100	0.750	0.885	0.884	0.678	0.896	0.897	0.711	0.874	0.876	0.688	0.885	0.888
50	50	0.683	0.891	0.888	0.675	0.880	0.877	0.690	0.881	0.879	0.678	0.885	0.891
50	100	0.714	0.886	0.885	0.661	0.883	0.883	0.691	0.879	0.889	0.668	0.864	0.872
Location-scale model ($\zeta = 0.5$) and $u \sim \chi^2_3$													
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.735	0.840	0.862	0.693	0.818	0.842	0.746	0.719	0.676	0.784	0.727	0.701
100	10	0.666	0.868	0.859	0.645	0.849	0.837	0.641	0.783	0.790	0.679	0.774	0.762
200	5	0.714	0.831	0.856	0.685	0.839	0.842	0.722	0.750	0.669	0.770	0.731	0.612
200	10	0.661	0.886	0.850	0.658	0.865	0.835	0.637	0.777	0.771	0.642	0.751	0.758
25	50	0.662	0.850	0.859	0.663	0.880	0.884	0.613	0.847	0.843	0.653	0.848	0.846
25	100	0.674	0.888	0.886	0.676	0.883	0.886	0.662	0.877	0.879	0.651	0.864	0.872
50	50	0.633	0.863	0.869	0.685	0.896	0.899	0.661	0.852	0.856	0.672	0.866	0.868
50	100	0.646	0.861	0.874	0.685	0.887	0.887	0.670	0.863	0.873	0.668	0.856	0.860

Table 4.2. *Empirical coverage probabilities of the bootstrap confidence interval for a nominal 90% level. See Table 4.1 for definitions.*

We generate data from several variations of the basic model. In one variant of the model, α_i is an i.i.d. Gaussian random variable. In another, we generate $\alpha_i = i/N$ for $1 \le i \le N$ as in Galvao, Gu, and Volgushev (2020). We use $\zeta \in \{0, 0.5\}$, and thus, $\beta_0 = 1$ in the location shift version of the model and $\beta_0 = 1 + 0.5F_u(\tau)^{-1}$ in the location-scale shift case. Lastly, we consider three different distributions for the error term. We assume that u_{it} is distributed as $\mathcal{N}(0,1)$, a t distribution with 3 degrees of freedom (t_3), or χ^2_3.

Tables 4.1, 4.2, 4.3, and 4.4 present coverage probabilities for a nominal 90% confidence interval for the slope parameter β_0. We present coverage probabilities using the empirical distribution of the bootstrap estimator (Tables 4.1 and 4.2), as well as coverage probabilities of the asymptotic Gaussian confidence interval (Tables 4.3 and 4.4). In the latter case, the coverage is constructed using the standard error of the corresponding bootstrap procedure. Tables 4.1 and 4.3 present results for the

		Quantile 0.5						Quantile 0.75					
		Method:			Method:			Method:			Method:		
N	T	CS	WB		CS	WB		CS	WB		CS	WB	
		PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE
		Location shift model ($\zeta = 0$) and $u \sim \mathcal{N}(0,1)$											
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.951	0.908	0.910	0.862	0.922	0.919	0.930	0.896	0.894	0.845	0.899	0.899
100	10	0.971	0.917	0.862	0.864	0.913	0.870	0.969	0.909	0.913	0.861	0.910	0.922
200	5	0.953	0.925	0.927	0.849	0.919	0.922	0.918	0.902	0.902	0.831	0.912	0.913
200	10	0.974	0.933	0.853	0.868	0.925	0.868	0.975	0.910	0.920	0.876	0.904	0.924
25	50	0.998	0.911	0.909	0.916	0.894	0.894	1.000	0.889	0.892	0.912	0.886	0.886
25	100	1.000	0.910	0.910	0.972	0.908	0.907	1.000	0.901	0.904	0.944	0.902	0.904
50	50	1.000	0.907	0.905	0.922	0.888	0.885	1.000	0.895	0.897	0.918	0.907	0.908
50	100	1.000	0.908	0.904	0.973	0.904	0.903	1.000	0.900	0.900	0.965	0.913	0.907
		Location shift model ($\zeta = 0$) and $u \sim t_3$											
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.945	0.922	0.918	0.860	0.920	0.916	0.917	0.887	0.885	0.814	0.865	0.901
100	10	0.969	0.923	0.873	0.856	0.923	0.878	0.964	0.908	0.916	0.857	0.894	0.907
200	5	0.946	0.940	0.939	0.826	0.913	0.931	0.911	0.898	0.901	0.774	0.841	0.903
200	10	0.967	0.920	0.822	0.850	0.933	0.838	0.960	0.901	0.915	0.848	0.880	0.897
25	50	0.997	0.926	0.921	0.924	0.912	0.906	0.989	0.890	0.889	0.901	0.890	0.889
25	100	1.000	0.895	0.893	0.950	0.893	0.892	0.999	0.879	0.880	0.935	0.902	0.902
50	50	0.998	0.904	0.899	0.900	0.887	0.881	0.999	0.895	0.897	0.910	0.906	0.909
50	100	1.000	0.896	0.891	0.955	0.902	0.898	0.998	0.890	0.890	0.917	0.875	0.874
		Location shift model ($\zeta = 0$) and $u \sim \chi^2_3$											
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.897	0.917	0.918	0.832	0.892	0.916	0.857	0.805	0.824	0.791	0.750	0.800
100	10	0.905	0.898	0.873	0.830	0.879	0.844	0.886	0.835	0.852	0.852	0.836	0.848
200	5	0.897	0.914	0.922	0.824	0.893	0.917	0.828	0.784	0.820	0.798	0.784	0.820
200	10	0.912	0.904	0.845	0.823	0.877	0.831	0.884	0.834	0.858	0.849	0.831	0.846
25	50	0.968	0.881	0.883	0.864	0.894	0.896	0.904	0.851	0.852	0.846	0.868	0.870
25	100	0.994	0.889	0.890	0.870	0.878	0.879	0.956	0.878	0.885	0.831	0.856	0.860
50	50	0.965	0.878	0.874	0.871	0.895	0.894	0.916	0.865	0.864	0.859	0.876	0.878
50	100	0.992	0.882	0.882	0.890	0.894	0.894	0.965	0.870	0.870	0.853	0.868	0.868

TABLE 4.3. *Empirical coverage probabilities of the asymptotic Gaussian confidence interval for a nominal 90% level. See Table 4.1 for definitions.*

location shift model ($\zeta = 0$), while Tables 4.2 and 4.4 present results for the location-scale shift model ($\zeta = 0.5$). The tables present results for $\tau \in \{0.50, 0.75\}$, based on different combinations of $N \in \{25, 50, 100, 200\}$ and $T \in \{5, 10, 50, 100\}$. The number of bootstrap repetitions is set to 400, and the results are obtained by using 1000 random samples.

The tables show results for two bootstrap methods. The cross-sectional pairs bootstrap (CS) samples over i with replacement, keeping the entire block of time series observations. The wild bootstrap (WB) is implemented as discussed in Section 2.3. We first obtain residuals $\hat{u}_{it}$ using the penalized quantile regression (2.6), which is labeled 'PQR' in the tables. The tuning parameter is obtained as $\hat{\lambda}_T = b_T \tilde{\lambda}$ where $\tilde{\lambda}$ is obtained by cross-validation and $b_T = 0.5T^{-\nu}$ controls the bias. The selection of $\nu = 1$ performed well in the simulations and it is consistent with Theorem 1. As in the case of the wild bootstrap estimator proposed by Feng, He, and Hu (2011), a finite sample correction is recommended.

		Quantile 0.5						Quantile 0.75					
		Method:			Method:			Method:			Method:		
N	T	CS	WB		CS	WB		CS	WB		CS	WB	
		PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE	PQR	PQR	FE
Location-scale shift model ($\zeta = 0.5$) and $u \sim \mathcal{N}(0,1)$													
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.851	0.893	0.886	0.863	0.909	0.905	0.812	0.841	0.837	0.830	0.876	0.843
100	10	0.833	0.912	0.865	0.821	0.910	0.860	0.819	0.873	0.871	0.825	0.887	0.877
200	5	0.838	0.899	0.896	0.834	0.906	0.895	0.784	0.838	0.831	0.810	0.884	0.837
200	10	0.833	0.917	0.845	0.817	0.913	0.847	0.812	0.873	0.870	0.801	0.884	0.879
25	50	0.897	0.903	0.903	0.837	0.909	0.905	0.883	0.894	0.893	0.813	0.873	0.874
25	100	0.938	0.912	0.909	0.827	0.902	0.901	0.922	0.901	0.900	0.824	0.898	0.898
50	50	0.887	0.901	0.897	0.801	0.896	0.886	0.882	0.890	0.891	0.821	0.888	0.890
50	100	0.943	0.909	0.906	0.824	0.904	0.900	0.919	0.891	0.888	0.838	0.907	0.907
Location-scale shift model ($\zeta = 0.5$) and $u \sim t_3$													
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.864	0.915	0.898	0.882	0.918	0.894	0.816	0.816	0.789	0.857	0.871	0.793
100	10	0.845	0.912	0.864	0.845	0.910	0.870	0.847	0.897	0.873	0.832	0.874	0.854
200	5	0.884	0.919	0.921	0.855	0.910	0.916	0.778	0.803	0.756	0.836	0.860	0.792
200	10	0.830	0.913	0.842	0.821	0.919	0.827	0.786	0.842	0.817	0.793	0.849	0.817
25	50	0.894	0.919	0.912	0.851	0.916	0.914	0.856	0.877	0.878	0.814	0.883	0.881
25	100	0.902	0.887	0.883	0.826	0.903	0.903	0.874	0.886	0.886	0.831	0.890	0.893
50	50	0.882	0.894	0.889	0.791	0.885	0.873	0.856	0.882	0.884	0.820	0.899	0.897
50	100	0.908	0.886	0.882	0.815	0.889	0.885	0.893	0.892	0.892	0.805	0.879	0.878
Location-scale model ($\zeta = 0.5$) and $u \sim \chi^2_3$													
		$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$			$\alpha_i \sim \mathcal{N}(0,1)$			$\alpha_i = i/N$		
100	5	0.857	0.887	0.868	0.829	0.857	0.847	0.881	0.818	0.672	0.896	0.843	0.714
100	10	0.837	0.886	0.847	0.822	0.871	0.832	0.821	0.830	0.786	0.822	0.830	0.779
200	5	0.833	0.881	0.858	0.823	0.860	0.844	0.875	0.850	0.641	0.884	0.842	0.584
200	10	0.845	0.903	0.835	0.814	0.883	0.821	0.815	0.823	0.778	0.799	0.818	0.758
25	50	0.803	0.858	0.858	0.815	0.890	0.886	0.804	0.851	0.850	0.806	0.859	0.857
25	100	0.847	0.893	0.900	0.819	0.885	0.885	0.829	0.881	0.884	0.808	0.863	0.868
50	50	0.819	0.874	0.872	0.836	0.907	0.907	0.808	0.859	0.861	0.826	0.878	0.876
50	100	0.827	0.876	0.877	0.829	0.895	0.892	0.825	0.875	0.878	0.811	0.866	0.868

TABLE 4.4. *Empirical coverage probabilities of the asymptotic Gaussian confidence interval for a nominal 90% level. See Table 4.1 for definitions.*

We adopt an adjustment following closely the `R` package `quantreg` by Koenker (2021). In our case, we adjust the residuals with the influence function and sign function following the Bahadur representation of the estimator derived in Theorem 2. Then, we generate $u^*_{it} = w_{it}|\hat{u}_{it}|$, where w_{it} is an i.i.d. random variable distributed as a two-point distribution with probabilities τ and $1-\tau$ at $w_{it} = -2\tau$ and $w_{it} = 2(1-\tau)$. Lastly, we generate the dependent variable as $y^*_{it} = \hat{\alpha}_i + \hat{\beta}x_{it} + u^*_{it}$. The performance of the estimator (2.7) was similar and the results are not presented here to save space. Finally, we include the estimator (2.6) defined for $\lambda_T = 0$ and it is labeled 'FE'.

Following the result presented in Theorem 3, the coverage probabilities in Table 4.1 are obtained considering the quantiles of the empirical distribution of $\sqrt{NT}(\beta^* - \hat{\beta})$. As can be seen in the upper block of Table 4.1, the performance of the WB bootstrap estimators are excellent, and they are in general around the specified coverage probability. Furthermore, performance improves with T, and

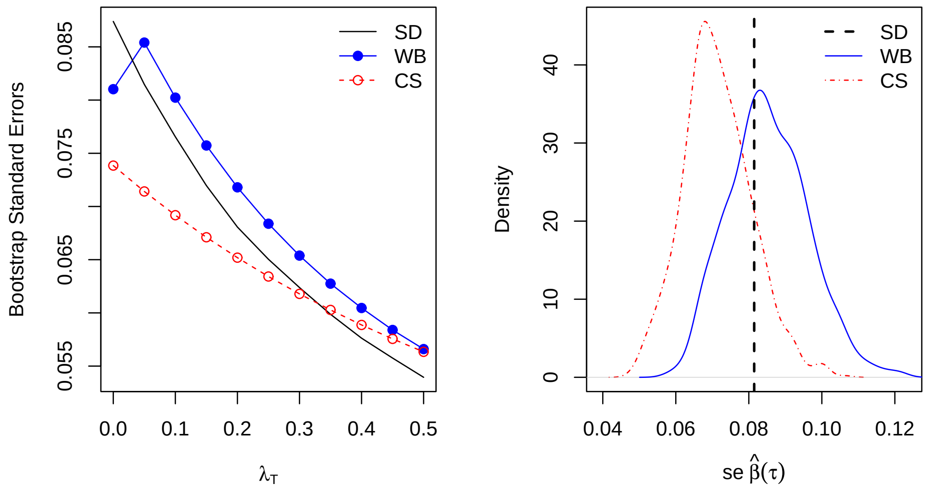

FIGURE 4.1. *The performance of the bootstrap estimators as λ_T increases. SD denotes standard deviation of the penalized estimator, CS denotes cross-sectional pair bootstrap, and WB denotes wild bootstrap estimator* (2.6)*.*

tends to be similar for both 0.5 and 0.75 quantiles. On the other hand, the performance of the CS estimator is poor, with estimates not approaching to specified nominal values. In the lower parts of the table, we present the performance of the estimators for different distributions F_u. The WB method continues to perform better than CS, and, as expected, the estimation of the higher quantile is more challenging in the χ^2_3 case. In all the variations of the model considered in the table, the WB estimator performs much better than the CS estimator.

The results for the location-scale shift model presented in Table 4.2 are similar. We continue to see that the WB bootstrap performs better than the CS method. This conclusion holds when we consider asymptotic Gaussian confidence intervals obtained using bootstrap standard errors se(β^*) (see Tables 4.3 and 4.4). Moreover, the tables confirm two results that were expected. First, as T increases relative to N, the coverage of the WB improves. Second, the performance of WB in the case of $\lambda_T = 0$ reveals that, in general, the procedure proposed in this paper is valid for approximating the distribution of the fixed effects estimator.

We finish the section by briefly documenting the relative performance of the estimators of the standard errors. We generate data from a location-scale shift model ($\zeta = 0.5$) when the error term $u_{it} \sim \mathcal{N}(0,1)$ and $\alpha_i \sim \mathcal{N}(0,1)$, by setting $N = 100$, $T = 10$, and $\tau = 0.5$. The left panel of Figure 4.1 shows CS and WB bootstrap estimates of the standard error, se(β^*), and the standard deviation of the penalized estimator, sd($\hat{\beta}$). The figure shows the advantage of the penalized estimator relative to the fixed effects estimator, as the standard deviation of the estimator is decreasing as λ_T increases. We also see that the WB procedure performs better than CS when λ_T is relatively small, and the performance of the WB estimator does not seem to change over the degree of shrinkage of the individual effects, as the bias appears to be roughly constant over λ_T. Using the right panel in Figure 4.1, we explore further the difference in performance between approaches. The empirical distribution obtained by the CS procedure is not centered at the true value, and the distribution of the standard error of the WB is centered at $\mathrm{se}(\hat{\beta}) = 0.081$ (with $\lambda_T = 0.05$).

5. Extensions

In this section, we investigate the consistency of the wild bootstrap under different conditions. First, we extend the results of Theorems 1 and 2 to allow for dependent data, and then we focus on the consistency of the wild bootstrap. In such case, we use the following assumptions:

C1. The processes $\{(y_{it}, \boldsymbol{x}_{it}), t \in 1, 2, \ldots\}$ are strictly stationary for each i and β-mixing, and independent across i. Letting $\{\beta_i(j)\}_j$ denote the β-mixing coefficients, assume that there are constants $0 < a < 1$ and $B > 0$ such that $\sup_i \beta_i(j) \leq Ba^j$ for all $j \geq 1$.

C2. The random vector (u_{it}, u_{it+j}) has a density conditional on $(\boldsymbol{x}_{it}, \boldsymbol{x}_{it+j})$ that is bounded uniformly over i and $j \geq 1$.

C3. Assume that the matrix $\boldsymbol{D}_N$ as defined in Assumption B5 exists and is positive definite for all N under Assumptions C1 and C2 and that $\boldsymbol{D} = \lim_{N\to\infty} \boldsymbol{D}_N$ exists and is positive definite. Also assume that

$$\tilde{\boldsymbol{V}} = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \mathrm{Var}\left(\sum_{t=1}^{T} (\tau - I(y_{it} < \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 + \alpha_{i0}))\left(\boldsymbol{x}_{it} - \varphi_i^{-1}\boldsymbol{E}_i\right) \right)$$

is positive definite.

Theorem 5 presents both consistency and asymptotic normality results for the estimator with dependent error terms.

Theorem 5. *Under Assumptions C1-C3, B3 and B4, if* $\log(N)^2/T \to 0$ *and* $\lambda_T = o_p(T)$ *as* $N, T \to \infty$*, the estimator* $\hat{\boldsymbol{\beta}}$ *is consistent. Moreover, if* $N^2(\log N)^3/T \to 0$ *and* $\lambda_T = o_p(T^{1/2}(\log N)^{1/2})$ *as* $N, T \to \infty$*, then*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \tilde{\boldsymbol{\Omega}}),$$

where $\tilde{\boldsymbol{\Omega}} = \boldsymbol{D}^{-1}\tilde{\boldsymbol{V}}\boldsymbol{D}^{-1}$.

Theorem 6 shows consistency of the bootstrap distribution estimator in the case of dependent errors. This more complex situation requires another assumption:

A4. Suppose that the weights w_{it} are independent over i and satisfy

$$\lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) \left(\mathrm{P}^*\{w_{it} < 0, w_{it+j} < 0\} - \mathrm{P}\left\{u_{it} \leq 0, u_{it+j} \leq 0 | \boldsymbol{x}_{it}, \boldsymbol{x}_{it+j}\right\}\right) = 0.$$

Assumption A4 is a high-level assumption on the distribution of bootstrap weights. The assumption guarantees that the variance of the bootstrap estimator is bounded and sufficiently close to the true variance, because the weights mimic the within-unit dependence structure of the errors. A feasible version could use a plug-in estimate of the average of the joint conditional CDFs of (u_{it}, u_{it+j}) to generate weights that satisfy the average probability. For example, let G_{W_j} denote the joint CDF of the bootstrap weights (w_{it}, w_{it+j}) for $1 \leq j \leq T-1$. Then, given residuals from estimation, we could construct a collection $\{G_{W_j}(0,0)\}_{j=1}^{T-1}$ that matches the sample analog of the condition from Assumption A4, that is, such that

$$\sum_{j=1}^{T-1}(1-j/T)G_{W_j}(0,0) = \sum_{j=1}^{T-1}(1-j/T)\frac{1}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j} I(\hat{u}_{it} \leq 0, \hat{u}_{it+j} \leq 0).$$

For example, we may construct correlated normal random variables using a simple parametric model to simulate a stationary sequence $\{z_{it}\}_{t=1}^{T}$ that is marginally standard normal and has known correlations $\{\rho_j\}_{j=1}^{T-1}$. Letting Φ be the standard normal CDF and G_W^{-1} be the quantile function for a distribution that satisfies Assumptions A1-A3, bootstrap weights defined by $w_{it} = G_W^{-1}(\Phi(z_{it}))$ have marginal distribution function G_W and $G_{W_j}(0,0) = \mathrm{P}\left\{z_{it} \leq \Phi^{-1}(\tau), z_{it+j} \leq \Phi^{-1}(\tau)\right\}$. Then we can tune the correlations to satisfy the equality given in the above display. Letting $\Phi_2(\cdot,\cdot,\rho)$ denote the bivariate normal CDF of two random variables with standard normal marginals and correlation ρ, we can solve numerically for $\{\hat{\rho}_j\}$ such that

$$\sum_{j=1}^{T-1}(1-j/T)\Phi_2(\Phi^{-1}(\tau),\Phi^{-1}(\tau),\hat{\rho}_j) = \sum_{j=1}^{T-1}(1-j/T)\frac{1}{N(T-j)}\sum_{i=1}^{N}\sum_{t=1}^{T-j} I(\hat{u}_{it} \leq 0, \hat{u}_{it+j} \leq 0).$$

To specify the example further, if we use an AR(1) model $z_{it} = rz_{it-1} + e_{it}$, $|r| < 1$ and $e_{it} \sim \mathcal{N}(0,1)$ to generate z_{it}, then $\rho_j = r^j$ for each j, and the solution $\hat{r}$ is the root of a univariate function. Furthermore, if $z_{it} = \hat{r}z_{it-1} + e_{it}$, then $z_{it}\sqrt{1-\hat{r}^2}$ is distributed as standard normal and the bootstrap weight is generated as $w_{it} = G_W^{-1}(\Phi(z_{it}\sqrt{1-\hat{r}^2}))$.

Theorem 6. *Suppose that the bootstrap weights satisfies assumptions A1-A4 and the data satisfy assumptions C1-C3, B3 and B4. If $N^2(\log N)^3/T \to 0$ and $\lambda_T = o_p(T^{1/2}(\log N)^{1/2})$ as $N, T \to \infty$, then*

$$\sup_{b\in\mathbb{R}^p}\left|\mathrm{P}\left\{\sqrt{NT}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) \leq b|\boldsymbol{S}\right\} - \mathrm{P}\left\{\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq b\right\}\right| \xrightarrow{p} 0,$$

where $\boldsymbol{S}$ denotes the observed sample and $\boldsymbol{\beta}^*$ denotes the slope estimator defined by (2.6).

Finally, we investigate if the conditions on the size of T relative to N needed for the asymptotic normality in Theorem 2 can be improved, especially in the light of recent work by Galvao, Gu, and Volgushev (2020). If instead of focusing on the stochastic order of the terms of the Bahadur representation of the penalized estimator, we focus on the expected values of the remainder terms, it is possible to show that the rates can be improved substantially. In order to show asymptotic normality, we employ the following assumption about the behavior of the penalty parameter.

B6. For some $\kappa \geq 2$, there exists a constant $K > 0$ such that $\mathrm{P}\left\{\lambda_T > KT^{1/2}(\log T)^{1/2}\right\} = O(T^{-\kappa})$.

Assumption B6 dictates the rate at which the probability of observing large a λ_T becomes small asymptotically. As illustrated in remark S.1, it is needed to provide a tail bound for the distribution of individual effects, which figure in the remainder terms of the Bahadur representation used to find the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ (such a bound holds naturally for terms related to minimizing the quantile regression objective function with bounded regressors, a fact used extensively in Galvao, Gu, and Volgushev (2020)). In the theorem below, we require $\lambda_T = O_p(\log T) = o_p(T^{1/2}(\log T)^{1/2})$, so this assumption only mildly strengthens the other regularity conditions.

Theorem 7. *Under Assumptions B1 and B3-B6, if* $N(\log T)^2/T \to 0$ *and* $\lambda_T = O_p(\log T)$ *as* $N, T \to \infty$*, then*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = \boldsymbol{D}^{-1}\boldsymbol{V}\boldsymbol{D}^{-1}$*.*

The proof in Theorem 7 uses an infeasible estimator $\tilde{\alpha}_i$ that is obtained considering T observations $y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0$. The difference between $\tilde{\alpha}_i$ and $\hat{\alpha}_i$ converges to zero as the slope coefficient $\hat{\boldsymbol{\beta}}$ converges in probability towards $\boldsymbol{\beta}_0$, under the condition on λ_T. Therefore, the remainder terms of the corresponding Bahadur representations are sufficiently close, leading to the improvements in the rates first obtained in Galvao, Gu, and Volgushev (2020) for the fixed effects estimator. We now show the consistency of the bootstrap distribution estimator under these relatively closer orders of N and T.

Theorem 8. *Under Assumptions A1-A3 and the conditions of Theorem 7,*

$$\sup_{b \in \mathbb{R}^p} \left| \mathrm{P}\left\{ \sqrt{NT}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) \leq b | \boldsymbol{S} \right\} - \mathrm{P}\left\{ \sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq b \right\} \right| \xrightarrow{p} 0.$$

where $\boldsymbol{S}$ *denotes the observed sample and* $\boldsymbol{\beta}^*$ *denotes the slope estimator defined by* (2.6)*.*

6. An Empirical Illustration

In recent years, policy makers and the general public have been debating and re-evaluating several aspects of trade, including the benefits of trade agreements (Burfisher, Robinson, and Thierfelder, 2001; Hakobyan and McLaren, 2016, among others). An important question is whether workers have been negatively affected by the North American Free Trade Agreement (NAFTA), which was signed by the governments of the United States of America, Canada, and Mexico in 1993. Hakobyan and McLaren (2016) find that the effect of NAFTA on *average* wage growth in the period 1990-2000 was negative. In this section, we use similar data and apply our approach to study the distributional impact of NAFTA. Our findings suggest that the agreement increased wage inequality. Low-wage workers experienced significant negative wage growth, while high-wage workers experienced, in general, significant positive wage growth. Our results are similar to evidence on the effect of Chinese imports on low-wage American workers (Chetverikov, Larsen, and Palmer, 2016).

6.1. Data.

Following Hakobyan and McLaren (2016), we use a 5% sample from the U.S. Census. We employ two cross-sectional samples in the year 1990 and 2000, and therefore, workers in the sample are observed once. The longitudinal nature of the analysis comes from exploiting the fact that we observe multiple individuals in a given industry and location. The sample includes workers between 25 and 64 years of age who reported positive income. We have demographic information including age, gender, marital status, race, and educational attainment of the worker classified in four categories: high school dropout, high school graduate, some college, and college graduate.

The data on U.S. tariffs and Mexico's revealed comparative advantage (RCA) are obtained from Hakobyan and McLaren (2016). Using their data, we have access to average U.S. tariffs by industry of employment of the worker and location (or Consistent Public-Use Microdata Area, abbreviated *conspuma*) of residence of the worker. In 1990, the average tariff by industry in 1990 was 2.1% percent (with a standard deviation of 3.9%), while the average local tariff by conspuma level was 1.03% (with a standard deviation of 0.67%). In the period 1990-2000, the tariffs decreased 1.7% at the industry level and 0.9% at the conspuma level. These descriptive statistics are used in the next section to estimate the percentage change in wages associated with the reduction in tariffs. We consider all industries with the exception of agriculture.

6.2. **Model.** To investigate the effect of NAFTA on the wages of American workers, we consider a specification that allows for the impact of the trade agreement to vary by industry, location, and educational attainment of the worker. To that end, we consider the following model as in Hakobyan and McLaren (2016):

$$y_{ijc} = \boldsymbol{\beta}_{1L}'\boldsymbol{L}_{ic} + \boldsymbol{\beta}_{2L}'\Delta\boldsymbol{L}_{ic} + \boldsymbol{\beta}_{1I}'\boldsymbol{I}_{ij} + \boldsymbol{\beta}_{2I}'\Delta\boldsymbol{I}_{ij} + \boldsymbol{X}_{ijc}'\boldsymbol{\Pi} + \alpha_{jc} + u_{ijc}, \tag{6.1}$$

where the response variable y_{ijc} is the logarithm of wages for worker i, who is employed in industry j and resides in conspuma c, $\boldsymbol{L}_{ic}$ and $\Delta\boldsymbol{L}_{ic}$ are location variables to be described below, $\boldsymbol{I}_{ij}$ and $\Delta\boldsymbol{I}_{ij}$ are industry variables, $\boldsymbol{X}_{ijc}$ is the vector of control variables considered in Hakobyan and McLaren (2016), and α_{jc} is a industry-conspuma effect. The error term is denoted by u_{ijc}.

The location variables are defined as $\boldsymbol{L}_{ic} = (L_{ic,1}, L_{ic,2}, L_{ic,3}, L_{ic,4})'$, where $L_{ic,k}$ is the product of an indicator for educational category k of worker i, an indicator variable for whether i is in the 2000 sample, and the average tariff in the conspuma of residence of worker i. Similarly, we can define $\Delta\boldsymbol{L}_{ic} = (\Delta L_{ic,1}, \Delta L_{ic,2}, \Delta L_{ic,3}, \Delta L_{ic,4})'$, as the change in $\boldsymbol{L}_{ic}$ due to the change in tariffs between 1990 and 2000 in the conspuma of residence of worker i. In terms of the industry variables, $\boldsymbol{I}_{ij} = (I_{ij,1}, I_{ij,2}, I_{ij,3}, I_{ij,4})'$, where $I_{ij,k}$ is the product of an indicator for educational category k, the RCA in industry j, an indicator variable for whether i is in the 2000 sample, and the tariff of the industry that employs worker i. Similarly, we define $\Delta\boldsymbol{I}_{ij} = (\Delta I_{ij,1}, \Delta I_{ij,2}, \Delta I_{ij,3}, \Delta I_{ij,4})'$, as the change in $\boldsymbol{I}_{ij,k}$ due to the tariff change between 1990 and 2000 in the industry that employs worker i.

Because industry latent factors and trends in some areas can affect wages and also the changes in tariffs, we employ the penalized estimator (2.2) to estimate a high-dimensional model with more than 84,000 parameters α_{jc}. The parameters of interest in equation (6.1) are $\boldsymbol{\beta}_{1L}$, $\boldsymbol{\beta}_{2L}$, $\boldsymbol{\beta}_{1I}$, and $\boldsymbol{\beta}_{2I}$, which measure the initial effect of tariffs by location and industry ($\boldsymbol{\beta}_{1L}$ and $\boldsymbol{\beta}_{1I}$), and the impact effect of a reduction of tariffs by location and industry ($\boldsymbol{\beta}_{2L}$ and $\boldsymbol{\beta}_{2I}$). Using these parameters, it is possible to obtain the effect of the trade agreement on wages. For instance, for locations that lost all of their protection after the introduction of NAFTA, the effect of the local average tariff is measured by $\boldsymbol{\beta}_{1L} - \boldsymbol{\beta}_{2L}$. Similarly, for industries that lost all of their protection, the effect of the industry tariff is $\boldsymbol{\beta}_{1I} - \boldsymbol{\beta}_{2I}$.

	Mean Effect	Quantiles 0.1	0.25	0.5	0.75	0.9
			High school dropouts			
Initial tariff effect, $\beta_{1I,1}$	2.018	1.156	2.603	1.880	0.991	0.434
	(1.274)	(0.820)	(1.120)	(1.047)	(0.847)	(1.105)
Impact effect, $\beta_{2I,1}$	3.569	3.082	4.625	3.245	1.666	0.600
	(1.544)	(0.945)	(1.314)	(1.191)	(1.024)	(1.290)
Industry effect: $\beta_{1I,1} - \beta_{2I,1}$	-1.551	-1.925	-2.022	-1.365	-0.675	-0.166
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.556]
			High school graduates			
Initial tariff effect, $\beta_{1I,2}$	1.081	5.015	2.224	0.426	-2.216	-2.933
	(0.870)	(0.523)	(0.626)	(0.747)	(0.515)	(0.436)
Impact effect, $\beta_{2I,2}$	2.315	9.259	4.318	1.337	-2.469	-3.855
	(1.086)	(0.595)	(0.736)	(0.873)	(0.618)	(0.543)
Industry effect: $\beta_{1I,2} - \beta_{2I,2}$	-1.234	-4.245	-2.094	-0.911	0.253	0.922
	[0.000]	[0.000]	[0.000]	[0.000]	[0.022]	[0.000]
			Some college			
Initial tariff effect, $\beta_{1I,3}$	-0.181	3.187	2.631	-0.921	-2.963	-3.765
	(1.146)	(0.820)	(1.172)	(1.151)	(0.779)	(0.879)
Impact effect, $\beta_{2I,3}$	1.070	7.360	4.889	-0.263	-3.452	-4.662
	(1.396)	(0.972)	(1.468)	(1.359)	(0.954)	(1.026)
Industry effect: $\beta_{1I,3} - \beta_{2I,3}$	-1.234	-4.245	-2.094	-0.911	0.253	0.922
	[0.000]	[0.000]	[0.000]	[0.000]	[0.022]	[0.000]
			College graduate			
Initial tariff effect, $\beta_{1I,4}$	-2.438	7.623	-1.363	-6.538	-7.681	-8.688
	(1.839)	(1.826)	(1.362)	(1.856)	(1.041)	(1.181)
Impact effect, $\beta_{2I,4}$	-2.095	12.840	-0.024	-8.066	-9.828	-11.490
	(2.175)	(2.215)	(1.630)	(2.291)	(1.178)	(1.301)
Industry effect: $\beta_{1I,4} - \beta_{2I,4}$	-0.343	-5.217	-1.339	1.528	2.147	2.801
	[0.439]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Location variables	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Number of α_{jc} effects	84,266	84,266	84,266	84,266	84,266	84,266
Observations	9,580,568	9,580,568	9,580,568	9,580,568	9,580,568	9,580,568

TABLE 6.1. *Regression results for the industry effects by educational category of the worker. We present standard errors in parenthesis, and p-values of a test for the equality of initial and impact effects in brackets.*

6.3. **Main empirical results.** Table 6.1 reports results for the coefficients $\boldsymbol{\beta}_{1I}$, and $\boldsymbol{\beta}_{2I}$ for the four educational categories. The table also shows results for $\beta_{1I,k} - \beta_{2I,k}$ for each educational category k and p-values (in brackets) of Wald-type tests for the null hypothesis $\mathrm{H}_0 : \beta_{1I,k} = \beta_{2I,k}$. The variance of the test is obtained using the proposed wild residual bootstrap procedure. The first column presents mean fixed effects regression results. The last five columns show penalized quantile regression (PQR) results with λ_T selected by cross-validation. The standard errors are obtained by the proposed wild bootstrap procedure. To save space, we do not present results on the control variables included in the vector $\boldsymbol{L}_{ic}$, $\Delta\boldsymbol{L}_{ic}$, and $\boldsymbol{X}_{ijc}$, but the results at the conditional mean are similar to the results in Table 4 (column (2)) in Hakobyan and McLaren (2016).

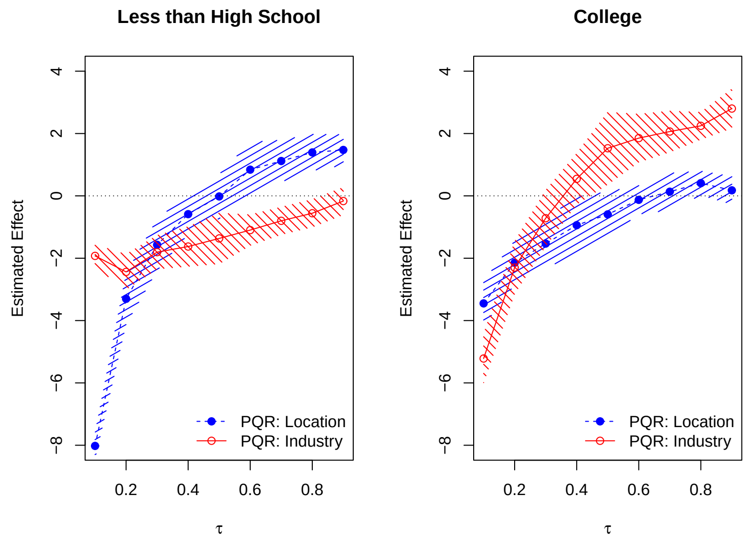

Figure 6.1. *Conditional wage growth impacts. PQR denotes penalized quantile regression and the dashed areas are 95% confidence intervals.*

Looking at the first set of estimates in the first rows, we see that an initial tariff estimate equal to 2.02 and an impact effect of 3.57. Based on the standard deviation of tariffs at the industry level, a 1% standard deviation increase in the initial industry tariff has an effect of reducing wages by $3.9\% \times -1.55$, or -6.05% in the period 1990-2000. This implies that, among industries with tariff declining after the introduction of NAFTA, average wage growth is negative for high school dropouts. The results, however, show that the average response does not summarize well the distributional impact of NAFTA. While the industry effect, which is measured as the difference between the initial effect and the impact effect, is negative (-1.93, or -7.50%) and significant for high school dropouts at the 0.1 quantile, it is small (-0.17, or -0.65%) and insignificant at the 0.9 quantile. Moreover, we find that the largest differences between the 0.1 and 0.9 effects are among college graduates in industries that lost all of their protection, suggesting that wage growth has been also unequal by educational attainment. Lastly, using Figure 6.1, we report point estimates and confidence intervals for the location and industry effects for high school dropouts and college graduates. The evidence reveals that inequality increased in the period after the implementation of the trade agreement.

7. Conclusion

In this article, we address the problem of estimating the distribution of the penalized quantile regression estimator for longitudinal data using a wild residual bootstrap procedure. Originally introduced by Koenker (2004) as a convenient alternative to the quantile regression estimator with fixed effects, the practical use of the penalized estimator has been limited by challenges involving inference. We show that the wild bootstrap procedure is asymptotically valid for approximating the distribution of the penalized estimator. We derive a series of new asymptotic results and carry out a simulation study that indicates that the wild residual bootstrap performs better than an alternative bootstrap approach commonly used in practice for similar estimators that do not include a penalty term.

Although the paper makes an important contribution by providing a valid method for statistical inference, there are several questions that remain to be answered. We believe that the procedure leads to valid inference in the case of J quantiles estimated simultaneously, but we leave this to future research. Moreover, under an assumption of sparsity as in other high-dimensional models, we expect changes in the consistency and asymptotic normality results. In terms of theoretical developments, we did not consider the case where α_i is a random effect. Lastly, the practical implementation of the wild bootstrap in the case of dependent data involves a few challenges. We hope to investigate these directions in future work.

Acknowledgements

We would like to thank the Co-editor Xiaohong Chen, an Associate Editor, and two referees for their detailed and constructive comments which have improved the paper considerably. We are also grateful to Antonio Galvao and Matt Harding for helpful comments and suggestions as well as seminar participants at the University of Kentucky, the 2019 CFE/CMStatistics conference, and the 2021 New York Camp Econometrics meeting.

Appendix A. Proof of main results

Remarks on notation and definitions: The estimators $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ depend on τ and λ_T, but we suppress this dependency for notational simplicity. The proofs refer to Knight's (1998) identity: $\rho_\tau(u-v)-\rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (I(u\le s) - I(u\le 0))\mathrm{d}s$, where $\rho_\tau = u(\tau - I(u<0))$ is the quantile regression check function and $\psi_\tau(u) = \tau - I(u<0)$ is the associated score function. Throughout the appendix, we define $\boldsymbol{\theta}_i = (\boldsymbol{\beta}', \alpha_i)'$ for each i, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$.

Proof of Theorem 1. Consistency follows from derivations analogous to those in Kato, Galvao, and Montes-Rojas (2012), tailored to accommodate a penalty term. Let $\hat{\boldsymbol{\theta}}$ be the minimizer of the normalized objective function

$$\mathbb{M}_{NT}(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i\right) + \frac{\lambda_T}{NT}\sum_{i=1}^{N}|\alpha_i|.$$

Define the i-th contribution to the objective function

$$\mathbb{M}_{Ti}(\boldsymbol{\theta}_i) := \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i\right) + \frac{\lambda_T}{T}|\alpha_i|$$

and let $\Delta_{Ti}(\boldsymbol{\theta}_i) = \mathbb{M}_{Ti}(\boldsymbol{\theta}_i) - \mathbb{M}_{Ti}(\boldsymbol{\theta}_{i0})$, that is,

$$\Delta_{Ti}(\boldsymbol{\theta}_i) = \frac{1}{T}\sum_{t=1}^{T}\left\{\rho_\tau\left(u_{it} - \boldsymbol{x}_{it}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0) - (\alpha_i - \alpha_{i0})\right) - \rho_\tau\left(u_{it}\right)\right\} + \frac{\lambda_T}{T}(|\alpha_i| - |\alpha_{i0}|).$$

By Knight's identity, $\Delta_{Ti}(\boldsymbol{\theta}_i) = \mathbb{V}_{Ti}^{(1)}(\boldsymbol{\theta}_i) + \mathbb{V}_{Ti}^{(2)}(\boldsymbol{\theta}_i)$, where

$$\mathbb{V}_{Ti}^{(1)}(\boldsymbol{\theta}_i) = -\frac{1}{T}\sum_{t=1}^{T}\left\{\boldsymbol{x}'_{it}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + (\alpha_i - \alpha_{i0})\right\}\psi_\tau(u_{it}) + \frac{\lambda_T}{T}(|\alpha_i| - |\alpha_{i0}|),$$

$$\mathbb{V}_{Ti}^{(2)}(\boldsymbol{\theta}_i) = \frac{1}{T}\sum_{t=1}^{T}\int_0^{\boldsymbol{x}'_{it}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+(\alpha_i-\alpha_{i0})} (I(u_{it}\le s) - I(u_{it}\le 0))\,\mathrm{d}s.$$

We first show the consistency of $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$. For each $\phi > 0$, define the ball $\mathcal{B}_i(\phi) := \{\boldsymbol{\theta}_i : \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i0}\|_1 \le \phi\}$ and the boundary $\partial\mathcal{B}_i(\phi) := \{\boldsymbol{\theta}_i : \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i0}\|_1 = \phi\}$. For each $\boldsymbol{\theta}_i \notin \mathcal{B}_i(\phi)$, define $\bar{\boldsymbol{\theta}}_i = r_i\boldsymbol{\theta}_i + (1-r_i)\boldsymbol{\theta}_{i0}$ where $r_i = \phi/\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i0}\|_1$. By construction, $r_i \in (0,1)$ and $\bar{\boldsymbol{\theta}}_i \in \partial\mathcal{B}_i(\phi)$.

Using the convexity of $\mathbb{M}_{Ti}(\boldsymbol{\theta}_i)$,

$$r_i\big(\mathbb{M}_{Ti}(\boldsymbol{\theta}_i) - \mathbb{M}_{Ti}(\boldsymbol{\theta}_{i0})\big) \ge \mathbb{M}_{Ti}(\bar{\boldsymbol{\theta}}_i) - \mathbb{M}_{Ti}(\boldsymbol{\theta}_{i0}) = \mathrm{E}\left[\Delta_{Ti}(\bar{\boldsymbol{\theta}}_i)\right] + \left(\Delta_{Ti}(\bar{\boldsymbol{\theta}}_i) - \mathrm{E}\left[\Delta_{Ti}(\bar{\boldsymbol{\theta}}_i)\right]\right). \tag{A.1}$$

Under Assumptions B1 and B2, we obtain, for $1 \le i \le N$,

$$\begin{aligned}\mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}_i)\right] &= \frac{\lambda_T}{T}(|\alpha_i| - |\alpha_{i0}|) + \mathrm{E}\left[\int_0^{\boldsymbol{x}'_{it}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+(\alpha_i-\alpha_{i0})} (F_i(s|\boldsymbol{x}_{i1}) - \tau)\mathrm{d}s\right] \\ &\ge \frac{\lambda_T}{T}(|\alpha_i| - |\alpha_{i0}|) + \epsilon_\phi\end{aligned}$$

for some $\epsilon_\phi > 0$. Using this in (A.1) results in

$$r_i\Delta_{Ti}(\boldsymbol{\theta}_i) \ge \epsilon_\phi + \frac{\lambda_T}{T}(|\bar{\alpha}_i| - |\alpha_{i0}|) + \left(\Delta_{Ti}(\bar{\boldsymbol{\theta}}_i) - \mathrm{E}\left[\Delta_{Ti}(\bar{\boldsymbol{\theta}}_i)\right]\right).$$

By the definition of $\hat{\boldsymbol{\theta}}_i$ as the minimizer of $N^{-1}\sum_i \mathbb{M}_{Ti}(\boldsymbol{\theta}_i)$, we have

$$\begin{aligned}\left\{\|\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}\|_1 > \phi\right\} &\subseteq \left\{\exists i \in \{1,\dots N\} : \hat{\boldsymbol{\theta}}_i \notin \mathcal{B}_i(\phi) \text{ and } \mathbb{M}_{Ti}(\hat{\boldsymbol{\theta}}_i) \le \mathbb{M}_{Ti}(\boldsymbol{\theta}_{i0})\right\} \\ &\subseteq \left\{\max_{1\le i\le N}\sup_{\boldsymbol{\theta}_i\in\mathcal{B}_i(\phi)}\Big|(\lambda_T/T)(|\alpha_i| - |\alpha_{i0}|) + \Delta_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}_i)\right]\Big| \ge \epsilon_\phi\right\}.\end{aligned}$$

Therefore, it is sufficient to show that

$$\lim_{N\to\infty}\mathrm{P}\left\{\max_{1\le i\le N}\sup_{\boldsymbol{\theta}_i\in\mathcal{B}_i(\phi)}\Big|(\lambda_T/T)(|\alpha_i| - |\alpha_{i0}|) + \Delta_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}_i)\right]\Big| \ge \epsilon_\phi\right\} = 0, \tag{A.2}$$

which is implied by

$$\max_{1\le i\le N}\mathrm{P}\left\{\sup_{\boldsymbol{\theta}_i\in\mathcal{B}_i(\phi)}\Big|(\lambda_T/T)(|\alpha_i| - |\alpha_{i0}|)\Big| + \Big|\Delta_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}_i)\right]\Big| \ge \epsilon_\phi\right\} = o(N^{-1}). \tag{A.3}$$

Normalize $\boldsymbol{\theta}_{i0} = \mathbf{0}_{p+1}$ for $1 \le i \le N$, so that $\mathcal{B}_i(\phi) = \mathcal{B}(\phi)$ for all $1 \le i \le N$. Let $h_{\boldsymbol{\theta}}(u, \boldsymbol{x}) := \rho_\tau(u - \boldsymbol{x}'\boldsymbol{\beta} - \alpha) - \rho_\tau(u) + (\lambda_T/T)|\alpha|$. By Assumption B3 and the reverse triangle inequality, letting $\Lambda = \lambda_U/T$, for some constant C,

$$|h_{\boldsymbol{\theta}}(u,\boldsymbol{x}) - h_{\boldsymbol{\theta}'}(u,\boldsymbol{x})| \le 2(1 + \|\boldsymbol{x}\| + \lambda_T/T)\left(\|\boldsymbol{\beta} - \boldsymbol{\beta}'\|_1 + |\alpha - \alpha'|\right) \le C(1 + M + \Lambda)\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_1.$$

For any $\phi > 0$, consider covering $\mathcal{B}(\phi)$, a compact set in $\mathbb{R}^{p+1}$, with L_1-balls of diameter ϵ over $B(\phi)$: generally $K = (\phi/\epsilon + 1)^{p+1}$ such balls are required. Cover $\mathcal{B}(\phi)$ with K balls of diameter $\epsilon/3\kappa$ where $\kappa = C(1 + M + \Lambda)$, and which have centers $\boldsymbol{\theta}^{(k)}$ for $k = 1, \ldots K$. Then the number of balls required is $K \leq \left(\frac{3\kappa\phi}{\epsilon} + 1\right)^{p+1} = O(\epsilon^{-(p+1)})$. Covering $B(\phi)$ with balls of this diameter implies that there is some $k \in \{1, \ldots K\}$ such that

$$\begin{aligned}
&\left|\Delta_{Ti}(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta})\right] - \Delta_{Ti}(\boldsymbol{\theta}^{(k)}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right]\right| \\
&\qquad \leq \left|\Delta_{Ti}(\boldsymbol{\theta}) - \Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right| + \left|\mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta})\right] - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right]\right| \leq 2\kappa\frac{\epsilon}{3\kappa} = \frac{2}{3}\epsilon.
\end{aligned}$$

Therefore for each $\boldsymbol{\theta} \in \mathcal{B}(\phi)$ there is a $k \in \{1, 2, ..., K\}$ such that

$$\left|\Delta_{Ti}(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta})\right]\right| \leq \left|\Delta_{Ti}(\boldsymbol{\theta}^{(k)}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right]\right| + \frac{2}{3}\epsilon,$$

and

$$\begin{aligned}
\mathrm{P}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{B}(\phi)}\left|\Delta_{Ti}(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta})\right]\right| > \epsilon\right\} &\leq \mathrm{P}\left\{\max_{1\leq k\leq K}\left|\Delta_{Ti}(\boldsymbol{\theta}^{(k)}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right]\right| + \frac{2\epsilon}{3} > \epsilon\right\} \\
&\leq \sum_{k=1}^{K}\mathrm{P}\left\{\left|\Delta_{Ti}(\boldsymbol{\theta}^{(k)}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right]\right| + \frac{2\epsilon}{3} > \epsilon\right\} \\
&= \sum_{k=1}^{K}\mathrm{P}\left\{\left|\Delta_{Ti}(\boldsymbol{\theta}^{(k)}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right]\right| > \epsilon/3\right\}.
\end{aligned}$$

For each term,

$$\begin{aligned}
\Delta_{Ti}(\boldsymbol{\theta}^{(k)}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right] &= \frac{1}{T}\sum_{t=1}^{T}\left(\rho_\tau(u_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}^{(k)} - \alpha^{(k)}) - \rho_\tau(u_{it})\right) \\
&\quad - \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\rho_\tau(u_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}^{(k)} - \alpha^{(k)}) - \rho_\tau(u_{it})\right)\right],
\end{aligned}$$

because the terms involving the penalty depend on $\alpha^{(k)}$ and cancel. Because each $\boldsymbol{\theta}^{(k)} \in \mathcal{B}(\phi)$, it can verified that $|\rho_\tau(u_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}^{(k)} - \alpha^{(k)}) - \rho_\tau(u_{it})| \leq (1 + M)\phi$. Hoeffding's inequality implies $\mathrm{P}\left\{\left|\Delta_{Ti}(\boldsymbol{\theta}^{(k)}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}^{(k)})\right]\right| > \epsilon/3\right\} \leq 2\exp\left\{-\frac{(\epsilon/3)^2T}{2(1+M)^2\phi^2}\right\}$. Therefore for any $\epsilon > 0$,

$$\mathrm{P}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{B}(\phi)}\left|\Delta_{Ti}(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta})\right]\right| > \epsilon/2\right\} \leq 2K\exp\{-DT\}.$$

Considering the penalty term, $(\lambda_T/T)(|\alpha_i| - |\alpha_{i0}|) \leq (\lambda_T/T)|\alpha_i - \alpha_{i0}| = O_p(\lambda_T/T)$, assuming $|\alpha_i - \alpha_{i0}| = O_p(1)$. Under the condition that $\lambda_T = o_p(T)$, $\lambda_T/T < \epsilon/2$ with probability increasing to 1. Therefore, consistency of $\hat{\boldsymbol{\beta}}$ is implied by the conditions $\log N = o(T)$ and $\lambda_T = o_p(T)$ as $N, T \to \infty$.

The consistency of $\hat{\boldsymbol{\beta}}$ implies consistency of $\hat{\alpha}_i$. Recall that $\hat{\alpha}_i = \arg\min \mathbb{M}_{NT}(\hat{\boldsymbol{\beta}}, \alpha)$. Isolating the part that depends on α_i, define the new ball $\mathcal{B}_i(\phi) := \{\alpha \in \mathbb{R} : |\alpha - \alpha_{i0}| \leq \phi\}$. For any α_i is not in $\mathcal{B}_i(\phi)$ define $\bar{\alpha}_i = r\alpha_i + (1 - r_i)\alpha_{i0}$ where $r_i = \phi/(|\alpha_i - \alpha_{i0}|)$ for $\phi > 0$. Because the objective function is convex

$$
\begin{aligned}
r_i\Big(\mathbb{M}_{Ti}(\hat{\boldsymbol{\beta}},\alpha_i) &- \mathbb{M}_{Ti}(\hat{\boldsymbol{\beta}},\alpha_{i0})\Big) \geq \mathbb{M}_{Ti}(\hat{\boldsymbol{\beta}},\bar{\alpha}_i) - \mathbb{M}_{Ti}(\hat{\boldsymbol{\beta}},\alpha_{i0}) \\
&= \{\mathbb{M}_{Ti}(\hat{\boldsymbol{\beta}},\bar{\alpha}_i) - \mathbb{M}_{Ti}(\boldsymbol{\beta}_0,\alpha_{i0})\} - \{\mathbb{M}_{Ti}(\hat{\boldsymbol{\beta}},\alpha_{i0}) - \mathbb{M}_{Ti}(\boldsymbol{\beta}_0,\alpha_{i0})\} \\
&= \Delta_{Ti}(\hat{\boldsymbol{\beta}},\bar{\alpha}_i) - \Delta_{Ti}(\hat{\boldsymbol{\beta}},\alpha_{i0}) \\
&= \{\Delta_{Ti}(\hat{\boldsymbol{\beta}},\bar{\alpha}_i) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\bar{\alpha}_i)\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} + \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\bar{\alpha}_i)\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \\
&\quad -\{\Delta_{Ti}(\hat{\boldsymbol{\beta}},\alpha_{i0}) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\alpha_{i0})\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\alpha_{i0})\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \\
&= \{\Delta_{Ti}(\hat{\boldsymbol{\beta}},\bar{\alpha}_i) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\bar{\alpha}_i)\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} - \{\Delta_{Ti}(\hat{\boldsymbol{\beta}},\alpha_{i0}) \\
&\quad -\mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\alpha_{i0})\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} + \{\mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\bar{\alpha}_i)\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta}_0,\bar{\alpha}_i)\right]\} \\
&\quad -\{\mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\alpha_{i0})\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta}_0,\alpha_{i0})\right]\} + \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta}_0,\bar{\alpha}_i)\right]
\end{aligned}
$$

Note that the last term $\mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta}_0,\bar{\alpha}_i)\right] \geq (\lambda_T/T)\mathrm{E}\left[|\bar{\alpha}_i| - |\alpha_{i0}|\right] + \epsilon_\phi$ for some $\epsilon_\phi > 0$ by Assumption B2. Thus, using similar calculations as before, we have

$$
\begin{aligned}
\big\{\exists i &\in \{1,\dots,N\} : |\hat{\alpha}_i - \alpha_{i0}| > \phi\big\} \\
&\subseteq \left\{\max_{1\leq i\leq N}\sup_{\alpha_i\in\mathcal{B}_i(\phi)}\Big((\lambda_T/T)||\alpha_i| - |\alpha_{i0}|| + \Big|\Delta_{Ti}(\hat{\boldsymbol{\beta}},\alpha) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\alpha)\right]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\Big|\Big) \geq \frac{\epsilon_\phi}{4}\right\} \\
&\cup \left\{\max_{1\leq i\leq N}\sup_{\alpha_i\in\mathcal{B}_i(\phi)}\Big|\mathrm{E}\Big[\Delta_{Ti}(\hat{\boldsymbol{\beta}},\alpha)\Big]\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta}_0,\alpha)\right]\Big| \geq \frac{\epsilon_\phi}{4}\right\} =: \mathcal{A}_{1N}\cup\mathcal{A}_{2N}.
\end{aligned}
$$

Because of the convexity of the objective function, the term involving $\bar{\alpha}_i$ is finite, and the entire bias term is $O_p(\lambda_T/T) = o_p(1)$ under the assumption on λ_T. By the consistency of $\hat{\boldsymbol{\beta}}$ and equation (A.3), $\mathrm{P}\{\mathcal{A}_{1N}\} \to 0$. Moreover, by B3 and the reverse triangle inequality, $|\mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta},\alpha)\right] - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\beta}_0,\alpha)\right]| \leq CM\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_1$ (due to cancellation of the penalty terms), $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \to 0$ implies $\mathrm{P}\{\mathcal{A}_{2N}\} \to 0$. □

Proof of Theorem 2. Define the scores with respect to $\boldsymbol{\beta}$ and α_i by

$$
\begin{aligned}
\mathbb{H}^{(\beta)}_{Ti}(\boldsymbol{\theta}_i) &:= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\psi_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i) \\
\mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\theta}_i) &:= \frac{1}{T}\sum_{t=1}^{T}\psi_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i) + \frac{\lambda_T}{T}\,\mathrm{sgn}(\alpha_i)
\end{aligned}
$$

and define $H^{(\beta)}_N(\boldsymbol{\theta}_i) := \mathrm{E}\left[\mathbb{H}^{(\beta)}_N(\boldsymbol{\theta}_i)\right]$ and $H^{(\alpha)}_{Ti}(\boldsymbol{\theta}_i) := \mathrm{E}\left[\mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\theta}_i)\right]$, that is,

$$
\begin{aligned}
H^{(\beta)}_{Ti}(\boldsymbol{\theta}_i) &= \mathrm{E}\left[\boldsymbol{x}_{i1}\left(\tau - F_i(\boldsymbol{x}_{i1}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + (\alpha_i - \alpha_{i0})|\boldsymbol{x}_{i1})\right]\right) \\
H^{(\alpha)}_{Ti}(\boldsymbol{\theta}_i) &= \mathrm{E}\left[\tau - F_i(\boldsymbol{x}_{i1}'(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + (\alpha_i - \alpha_{i0})|\boldsymbol{x}_{i1})\right] + (\lambda_T/T)\,\mathrm{sgn}(\alpha_i).
\end{aligned}
$$

First, derive a Bahadur representation for $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. For each i write

$$
\mathbb{H}^{(\alpha)}_{Ti}(\hat{\boldsymbol{\theta}}_i) = \mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\theta}_{i0}) + \Big(\mathbb{H}^{(\alpha)}_{Ti}(\hat{\boldsymbol{\theta}}_i) - H^{(\alpha)}_{Ti}(\hat{\boldsymbol{\theta}}_i) - \mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\theta}_{i0}) + H^{(\alpha)}_{Ti}(\boldsymbol{\theta}_{i0})\Big) + H^{(\alpha)}_{Ti}(\hat{\boldsymbol{\theta}}_i) - H^{(\alpha)}_{Ti}(\boldsymbol{\theta}_{i0}). \quad \text{(A.4)}
$$

Recalling the definitions made in Assumption B5 and the bounds in Assumption B4, expand the differentiable part of $H_{Ti}^{(\alpha)}$ around $\boldsymbol{\theta}_{i0}$ to find

$$H_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - H_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) = -\boldsymbol{E}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - \varphi_i(\hat{\alpha}_i - \alpha_{i0}) \\ + O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right) + O_p\left((\hat{\alpha}_i - \alpha_{i0})^2\right) + (\lambda_T/T)\left(\mathrm{sgn}(\hat{\alpha}_i) - \mathrm{sgn}(\alpha_{i0})\right).$$

Using the last expression and equation (A.4), solve for $\hat{\alpha}_i - \alpha_{i0}$ to find

$$\hat{\alpha}_i - \alpha_{i0} = -\varphi_i^{-1}\boldsymbol{E}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \varphi_i^{-1}\left(\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\alpha_{i0})\right) \\ + \varphi_i^{-1}\left(\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - H_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) + H_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})\right) \\ - \varphi_i^{-1}\left(\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\hat{\alpha}_i)\right) + O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right) + O_p\left((\hat{\alpha}_i - \alpha_{i0})^2\right). \quad \text{(A.5)}$$

Similarly, expand $H_{Ti}^{(\beta)}$ around $\boldsymbol{\theta}_{i0}$, noting $H_{Ti}^{(\beta)}(\boldsymbol{\theta}_{i0}) = \mathbf{0}_p$, to find

$$H_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) = -\boldsymbol{J}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - \boldsymbol{E}_i(\hat{\alpha}_i - \alpha_{i0}) + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) + O_p\left((\hat{\alpha}_i - \alpha_{i0})^2\right). \quad \text{(A.6)}$$

Substituting (A.5) in equation (A.6), after simplification, we obtain

$$H_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) = -(\boldsymbol{J}_i - \varphi_i^{-1}\boldsymbol{E}_i\boldsymbol{E}_i')(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - \varphi_i^{-1}\boldsymbol{E}_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\alpha_{i0})\right) \\ - \varphi_i^{-1}\boldsymbol{E}_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - H_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) + H_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})\right) \\ + \varphi_i^{-1}\boldsymbol{E}_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\hat{\alpha}_i)\right) + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|) + O_p\left((\hat{\alpha}_i - \alpha_{i0})^2\right) \quad \text{(A.7)}$$

Once again, for each i we have

$$\mathbb{H}_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) = \mathbb{H}_{Ti}^{(\beta)}(\boldsymbol{\theta}_{i0}) + \left(\mathbb{H}_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) - H_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{H}_{Ti}^{(\beta)}(\boldsymbol{\theta}_{i0})\right) + H_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i). \quad \text{(A.8)}$$

Substitute (A.8) into the left-hand side of (A.7) and solve for $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. Rearrange to find

$$(\boldsymbol{J}_i - \varphi_i^{-1}\boldsymbol{E}_i\boldsymbol{E}_i')(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|) = -\varphi_i^{-1}\boldsymbol{E}_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\alpha_{i0})\right) + \mathbb{H}_{Ti}^{(\beta)}(\boldsymbol{\theta}_{i0}) \\ - \varphi_i^{-1}\boldsymbol{E}_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - H_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) + H_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})\right) \\ + \left(\mathbb{H}_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) - H_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{H}_{Ti}^{(\beta)}(\boldsymbol{\theta}_{i0}) + H_{Ti}^{(\beta)}(\boldsymbol{\theta}_{i0})\right) \\ + \varphi_i^{-1}\boldsymbol{E}_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\hat{\alpha}_i)\right) - \mathbb{H}_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) + O_p\left((\hat{\alpha}_i - \alpha_{i0})^2\right). \quad \text{(A.9)}$$

It can be verified that for each i, $\left|\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i)\right| \leq 1/T$ (for $\lambda_T < \min\{\tau, 1-\tau\}T$). That implies the i-th individual effect estimate $\hat{\alpha}_i$ is between the $(\tau - (\lambda_T + 1)/T)$-th and $(\tau + (\lambda_T + 1)/T)$-th sample quantiles of the unit-i observations $\{y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}\}_{t=1}^{T}$. Therefore

$$\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\operatorname{sgn}(\hat{\alpha}_i) = \frac{1}{T}\sum_{t=1}^{T}\Big(\tau - I(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}} \le \hat{\alpha}_i)\Big) = O_p(\lambda_T/T). \tag{A.10}$$

Similarly, $\mathbb{H}_{Ti}^{(\beta)}(\hat{\boldsymbol{\theta}}_i) = O_p(\lambda_T/T)$. Now define $\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_i) = \mathbb{H}_{Ti}^{(\beta)}(\boldsymbol{\theta}_i) - \varphi_i^{-1}\boldsymbol{E}_i(\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_i) - (\lambda_T/T)\operatorname{sgn}(\alpha_i))$, $K_{Ti}^{(\theta)}(\boldsymbol{\theta}_i) = \mathrm{E}\left[\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_i)\right]$ and $\boldsymbol{D}_N = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{J}_i - \varphi_i^{-1}\boldsymbol{E}_i\boldsymbol{E}_i')$. Averaging equation (A.9) over i and using the above definitions and (A.10) we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + o_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|\right) &= \boldsymbol{D}_N^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0}) \\
&\quad + \boldsymbol{D}_N^{-1}\frac{1}{N}\sum_{i=1}^{N}\Big(\mathbb{K}_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - K_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0}) + K_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})\Big) \\
&\quad + O_p\left(\lambda_T/T\right) + O_p\left(\sup_i(\hat{\alpha}_i - \alpha_{i0})^2\right).
\end{aligned} \tag{A.11}$$

Next, we establish the rates of convergence for the estimators. Step 2 of the proof of Theorem 3.2 of Kato, Galvao, and Montes-Rojas (2012) shows that if $\sup_i|\hat{\alpha}_i - \alpha_{i0}| \vee \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(\delta_N)$, then for $d_{NT} = (|\log\delta_N|/T) \vee (\delta_N|\log\delta_N|/T)^{1/2}$,

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbb{K}_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - K_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0}) + K_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})\right\| = O_p(d_{NT}) = o_p(T^{-1/2}), \tag{A.12}$$

where the second equality follows from the consistency of $\hat{\boldsymbol{\theta}}$. The first term on the right-hand side of (A.11) is $O_p((NT)^{-1/2}) = o_p(T^{-1/2})$. Then, we have

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = o_p(T^{-1/2}) + O_p(T^{-1}\lambda_T) + O_p\left(\sup_i(\hat{\alpha}_i - \alpha_{i0})^2\right). \tag{A.13}$$

Using (A.13), we find that with probability approaching 1, there is some K such that

$$\begin{aligned}
\sup_i|\hat{\alpha}_i - \alpha_{i0}| &\le K\sup_i\left|\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) - \frac{\lambda_T}{T}\operatorname{sgn}(\alpha_{i0})\right| \\
&\quad + K\sup_i\left\|\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\alpha}}_i) - H_{Ti}^{(\alpha)}(\hat{\boldsymbol{\alpha}}_i) - \mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\alpha}_{i0}) + H_{Ti}^{(\alpha)}(\boldsymbol{\alpha}_{i0})\right\| + O_p(T^{-1}\lambda_T) + o_p(T^{-1/2}).
\end{aligned}$$

The first term in the above sum is mean zero and bounded. Hoeffding's inequality implies that for some K,

$$\begin{aligned}
&\mathrm{P}\Big\{\sup_i\left|\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) - \frac{\lambda_T}{T}\operatorname{sgn}(\alpha_{i0})\right| > T^{-1/2}(\log N)^{1/2}K\Big\} \\
&\qquad \le \sum_{i=1}^{N}\mathrm{P}\left\{\left|\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) - \frac{\lambda_T}{T}\operatorname{sgn}(\alpha_{i0})\right| > T^{-1/2}(\log N)^{1/2}K\right\} \le 2N^{1-K^2/2},
\end{aligned}$$

so that $\sup_i |\mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\theta}_{i0}) - (\lambda_T/T)\,\mathrm{sgn}(\alpha_{i0})| = O_p(T^{-1/2}(\log N)^{1/2})$. Step 3 of the proof of Theorem 3.2 of Kato, Galvao, and Montes-Rojas (2012) implies that

$$\sup_i \left\| \mathbb{H}^{(\alpha)}_{Ti}(\hat{\boldsymbol{\alpha}}_i) - H^{(\alpha)}_{Ti}(\hat{\boldsymbol{\alpha}}_i) - \mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\alpha}_{i0}) + H^{(\alpha)}_{Ti}(\boldsymbol{\alpha}_{i0}) \right\| = o_p(T^{-1/2}(\log N)^{1/2}).$$

Together, these estimates imply that if $\lambda_T = o_p(T^{1/2}(\log N)^{1/2})$, then

$$\sup_i |\hat{\alpha}_i - \alpha_{i0}| = O_p\left(T^{-1/2}(\log N)^{1/2}\right), \tag{A.14}$$

and via (A.13) that

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = o_p\left(T^{-1/2}(\log N)^{1/2}\right). \tag{A.15}$$

The condition on λ_T and the argument of Kato, Galvao, and Montes-Rojas (2012) implies that if $T^{-1}N^2(\log N)^3 \to 0$, we may rewrite equation (A.11) as

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{D}_N^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} (\boldsymbol{x}_{it} - \varphi_i^{-1}\boldsymbol{E}_i)\psi_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 - \alpha_{i0}) + o_p(1),$$

and the Lyapunov Central Limit Theorem implies that $\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. □

Proof of Theorem 3. In this proof, the notation $\boldsymbol{X}^* \overset{p^*}{\to} \boldsymbol{X}$ denotes convergence in probability of $\boldsymbol{X}^*$ to $\boldsymbol{X}$ under the resampling distribution, conditional on the observed sample $\boldsymbol{S}$. Similarly, let $\mathrm{E}^*[\cdot] = \mathrm{E}[\cdot|\boldsymbol{S}]$ and $\mathrm{P}^*\{\cdot\} = P\{\cdot|\boldsymbol{S}\}$ denote the expected value operator and probability calculated conditional on the data, and stochastic order symbols $O_{p^*}(\cdot)$ and $o_{p^*}(\cdot)$ are interpreted conditional on the observed sample. The proof is divided in two parts. The first part of the proof shows consistency by demonstrating that feasible and infeasible versions of the wild residual bootstrap estimator are equivalent as N and $T \to \infty$. The second part of the proof establishes asymptotic normality of $\boldsymbol{\theta}^*$.

For all i and t let $y^*_{it} = \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}} + \hat{\alpha}_i + w_{it}|\hat{u}_{it}|$, and let $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^{*\prime}, \boldsymbol{\alpha}^{*\prime})'$ be the solution of $\min_{\boldsymbol{\theta}} \mathbb{M}^*_{NT}(\boldsymbol{\theta})$ where

$$\mathbb{M}^*_{NT}(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau\left(y^*_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i\right) + \frac{\lambda_T}{NT}\sum_{i=1}^{N}|\alpha_i|.$$

Also define the i-th contribution to the objective function by $\mathbb{M}^*_{Ti}$,

$$\mathbb{M}^*_{Ti}(\boldsymbol{\theta}_i) = \frac{1}{T}\sum_{t=1}^{T} \rho_\tau\left(y^*_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i\right) + \frac{\lambda_T}{T}|\alpha_i|,$$

so that $\mathbb{M}^*_{NT} = \frac{1}{N}\sum_i \mathbb{M}^*_{Ti}$. Before examining $\mathbb{M}^*_{NT}$ and $\boldsymbol{\theta}^*$, consider an infeasible resampled objective function using the true error terms instead of the estimated residuals: let $y^\circ_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 + \alpha_{i0} + w_{it}|u_{it}|$ and define

$$\mathbb{M}^\circ_{Ti}(\boldsymbol{\theta}_i) = \frac{1}{T}\sum_{t=1}^{T} \rho_\tau\left(y^\circ_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i\right) + \frac{\lambda_T}{T}|\alpha_i|.$$

Let $\boldsymbol{\theta}^\circ$ be the minimizer of $\frac{1}{N}\sum_i \mathbb{M}^\circ_{Ti}(\boldsymbol{\theta}_i)$. As in the proof of Theorem 1, we define $\Delta^\circ_{Ti}(\boldsymbol{\theta}_i) = \mathbb{M}^\circ_{Ti}(\boldsymbol{\theta}_i) - \mathbb{M}^\circ_{Ti}(\boldsymbol{\theta}_{i0})$. Note that $\mathrm{E}^*\left[\Delta^\circ_{Ti}(\boldsymbol{\theta}_i)\right]$ is minimized at $\boldsymbol{\theta}_{i0}$.

Define the ϕ ball $\mathcal{B}_i(\phi) := \{\boldsymbol{\theta}_i : \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i0}\|_1 \leq \phi\}$ around $\boldsymbol{\theta}_{i0}$ and for $\boldsymbol{\theta}_i$ outside of the ball, define the weight $r_i = \phi/\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i0}\|_1$ and midpoint $\bar{\boldsymbol{\theta}}_i = r_i\boldsymbol{\theta}_i + (1-r_i)\boldsymbol{\theta}_{i0}$. Then

$$r_i(\mathbb{M}^\circ_{Ti}(\boldsymbol{\theta}_i) - \mathbb{M}^\circ_{Ti}(\boldsymbol{\theta}_{i0}) \geq \mathrm{E}^*\left[\Delta^\circ_{Ti}(\bar{\boldsymbol{\theta}}_i)\right] + \left(\Delta^\circ_{Ti}(\bar{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\Delta^\circ_{Ti}(\bar{\boldsymbol{\theta}}_i)\right]\right),$$

Similarly to the consistency proof, we have

$$\begin{aligned}\mathrm{E}^*\left[\Delta^\circ_{Ti}(\boldsymbol{\theta}_i)\right] =& \frac{\lambda_T}{T}\left\{|\alpha_i| - |\alpha_{i0}|\right\} \\ &+ \mathrm{E}^*\left[\frac{1}{T}\sum_{t=1}^T \int_0^{l_{it}(\boldsymbol{\theta}_i)} (I(w_{it}|u_{it}| \leq s) - I(w_{it}|u_{it}| \leq 0))\mathrm{d}s\right] \end{aligned} \tag{A.16}$$

where $l_{it}(\boldsymbol{\theta}_i) = \boldsymbol{x}_{it}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + (\alpha_i - \alpha_{i0})$. By Lemma S.2, equation (A.16) can be rewritten

$$\mathrm{E}^*\left[\Delta^\circ_{Ti}(\boldsymbol{\theta}_i)\right] = \frac{\lambda_T}{T}\left\{|\alpha_i| - |\alpha_{i0}|\right\} + \frac{1}{T}\sum_{t=1}^T f_i(0|\boldsymbol{x}_{it}) l_{it}(\boldsymbol{\theta}_i)' l_{it}(\boldsymbol{\theta}_i) + o_p(\sup_t \|l_{it}(\boldsymbol{\theta}_i)\|^2).$$

For $\bar{\boldsymbol{\theta}}_i$ on the ϕ ball around θ_{i0}, there is some $\epsilon_\phi > 0$ such that

$$\mathrm{E}^*\left[\Delta^\circ_{Ti}(\bar{\boldsymbol{\theta}}_i)\right] \geq \frac{\lambda_T}{T}\left\{|\bar{\alpha}_i| - |\alpha_{i0}|\right\} + \epsilon_\phi$$

where $\epsilon_\phi > 0$ and

$$r_i\Delta^\circ_{Ti}(\boldsymbol{\theta}_i) \geq \epsilon_\phi + \frac{\lambda_T}{T}(|\bar{\alpha}_i| - |\alpha_{i0}|) + \left(\Delta^\circ_{Ti}(\bar{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\Delta^\circ_{Ti}(\bar{\boldsymbol{\theta}}_i)\right]\right). \tag{A.17}$$

Then, similarly to the proof of consistency of $\hat{\boldsymbol{\theta}}$, the minimizer $\boldsymbol{\theta}^\circ$ is consistent if the following probability is satisfied:

$$\sup_{1\leq i\leq N} \mathrm{P}^*\left\{\sup_{\boldsymbol{\theta}_i, \in\mathcal{B}_i(\phi)} \left|(\lambda/T)(|\alpha_i| - |\alpha_{i0}|)\right| + \left|\Delta^\circ_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}^*\left[\Delta^\circ_{Ti}(\boldsymbol{\theta}_i)\right]\right| \geq \epsilon_\phi\right\} = o_p(N^{-1}). \tag{A.18}$$

The steps to show that $\boldsymbol{\theta}^\circ \xrightarrow{p^*} \boldsymbol{\theta}_0$ from this point on are identical to those in Theorem 1.

Having established the consistency of the infeasible estimator $\boldsymbol{\theta}^\circ$, denoting

$$\Delta^*_{Ti}(\boldsymbol{\theta}_i) = \mathbb{M}^*_{Ti}(\boldsymbol{\theta}_i) - \mathbb{M}^*_{Ti}(\hat{\boldsymbol{\theta}}_i) = \frac{1}{T}\sum_{t=1}^T \left(\rho_\tau(y^*_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i) - \rho_\tau(y^*_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}} - \hat{\alpha}_i)\right) + \frac{\lambda_T}{T}(|\alpha_i| - |\hat{\alpha}_i|),$$

we consider $\sup_{\boldsymbol{\theta}_i\in\Theta} |\Delta^*_{Ti}(\boldsymbol{\theta}_i) - \Delta^\circ_{Ti}(\boldsymbol{\theta}_i)|$. Notice that for each i,

$$\begin{aligned}\sup_{\boldsymbol{\theta}_i\in\mathbb{R}^{p+1}} |\Delta^*_{Ti}(\boldsymbol{\theta}_i) - \Delta^\circ_{Ti}(\boldsymbol{\theta}_i)| = &\Bigg|\frac{1}{T}\sum_{t=1}^T \Big(\rho_\tau\left(w_{it}|\hat{u}_{it}| - \boldsymbol{x}_{it}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - (\alpha_i - \hat{\alpha}_i)\right) \\ &- \rho_\tau\left(w_{it}|u_{it}| - \boldsymbol{x}_{it}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - (\alpha_i - \alpha_{i0})\right) \\ &- (\rho_\tau(w_{it}|\hat{u}_{it}|) - \rho_\tau(w_{it}|u_{it}|))\Big) + \frac{\lambda_T}{T}\left(|\hat{\alpha}_i| - |\alpha_{i0}|\right)\Bigg| \\ \leq M\left(1 + \frac{2}{T}\sum_{t=1}^T |w_{it}|\right)&\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| + \left(1 + \frac{\lambda_T}{T} + \frac{2}{T}\sum_{t=1}^T |w_{it}|\right)|\hat{\alpha}_i - \alpha_{i0}|. \end{aligned} \tag{A.19}$$

Using the consistency of $\hat{\boldsymbol{\theta}}_i$ for all i and as long as $\lambda_T/T = O_p(1)$, the average of these differences over i is $o_{p^*}(1)$ as $N, T \to \infty$ and also

$$\sup_{\boldsymbol{\theta}_i \in \mathbb{R}^{p+1}} |\Delta^*_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}^*\left[\Delta^*_{Ti}(\boldsymbol{\theta}_i)\right] - \{\Delta^\circ_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}^*\left[\Delta^\circ_{Ti}(\boldsymbol{\theta}_i)\right]\}| = o_{p^*}(1)$$

as $N, T \to \infty$. Finally, replacing the Δ° terms with Δ^* terms in (A.17) and (A.18) and approximating the inequalities with Δ° terms implies that $\boldsymbol{\theta}^* \stackrel{p^*}{\to} \boldsymbol{\theta}^\circ$. Therefore, the wild residual bootstrap estimator $\boldsymbol{\theta}^*$ is consistent because, as demonstrated above, $\boldsymbol{\theta}^\circ \stackrel{p^*}{\to} \boldsymbol{\theta}_0$.

Next consider the weak convergence of the estimator. Define the i-th contribution to the scores for $\mathbb{M}^*_{NT}$ with respect to $\boldsymbol{\beta}$ and α_i,

$$\mathbb{H}^{(\beta)*}_{Ti}(\boldsymbol{\theta}_i) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}\psi_\tau(u^*_{it} - \boldsymbol{x}'_{it}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - (\alpha_i - \hat{\alpha}_i))$$

and

$$\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}_i) = \frac{1}{T}\sum_{t=1}^{T} \psi_\tau(u^*_{it} - \boldsymbol{x}'_{it}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - (\alpha_i - \hat{\alpha}_i)) + \frac{\lambda_T}{T}\,\mathrm{sgn}(\alpha_i),$$

where $u^*_{it} = w_{it}|\hat{u}_{it}|$.

Write

$$\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) = \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) + \left(\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) - \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) - \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i)\right]\right) + \mathrm{E}^*\left[\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) - \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i)\right] \quad \text{(A.20)}$$

For the next part, make the following definitions, which are sample analogs to quantities defined in Assumption B5. Let $\bar{\varphi}_i = \frac{1}{T}\sum_t f_i(0|\boldsymbol{x}_{it})$, $\bar{\boldsymbol{E}}_i = \frac{1}{T}\sum_t f_i(0|\boldsymbol{x}_{it})\boldsymbol{x}_{it}$, $\bar{\boldsymbol{J}}_i = \frac{1}{T}\sum_t f_i(0|\boldsymbol{x}_{it})\boldsymbol{x}_{it}\boldsymbol{x}'_{it}$ and $\bar{\boldsymbol{D}}_N = \frac{1}{N}\sum_i(\bar{\boldsymbol{J}}_i - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i\bar{\boldsymbol{E}}'_i)$.

Part 2 of Lemma S.2 and $\boldsymbol{\theta}^*_i \stackrel{p^*}{\to} \hat{\boldsymbol{\theta}}_i$ imply that for all $1 \le i \le N$,

$$\mathrm{E}^*\left[\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) - \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i)\right] = -\frac{1}{T}\sum_{t=1}^{T} f_i(0|\boldsymbol{x}_{it})\left(\boldsymbol{x}'_{it}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) + (\alpha^*_i - \hat{\alpha}_i)\right) + \frac{\lambda_T}{T}(\mathrm{sgn}(\alpha^*_i) - \mathrm{sgn}(\hat{\alpha}_i)) + O_{p^*}\left((\alpha^*_i - \hat{\alpha}_i)^2 \vee \|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|^2\right) + O_p\left((\hat{\alpha}_i - \alpha_{i0})^2 \vee \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right). \quad \text{(A.21)}$$

Rewrite (A.20) using the above equation as

$$\begin{aligned}\alpha^*_i - \hat{\alpha}_i = {} & -\bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}'_i(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) + \bar{\varphi}_i^{-1}\left(\mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\hat{\alpha}_i)\right) \\ & + \bar{\varphi}_i^{-1}\left(\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) - \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) - \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i)\right]\right) \\ & - \bar{\varphi}_i^{-1}\left(\mathbb{H}^{(\alpha)*}_{Ti}(\boldsymbol{\theta}^*_i) - \frac{\lambda_T}{T}\,\mathrm{sgn}(\alpha^*_i)\right) + O_{p^*}\left((\alpha^*_i - \hat{\alpha}_i)^2 \vee \|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|^2\right) + O_p\left((\hat{\alpha}_i - \alpha_{i0})^2 \vee \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2\right).\end{aligned} \quad \text{(A.22)}$$

Similarly,

$$\mathbb{H}_{Ti}^{(\beta)*}(\boldsymbol{\theta}_i^*) = \mathbb{H}_{Ti}^{(\beta)*}(\hat{\boldsymbol{\theta}}_i) + \left(\mathbb{H}_{Ti}^{(\beta)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\beta)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\beta)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\beta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right) + \mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\beta)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\beta)*}(\hat{\boldsymbol{\theta}}_i)\right]. \quad \text{(A.23)}$$

Lemma S.2 can be used again to calculate the estimate

$$\mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\beta)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\beta)*}(\hat{\boldsymbol{\theta}}_i)\right] = -\bar{\boldsymbol{J}}_i(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) - \bar{\boldsymbol{E}}_i(\alpha_i^* - \hat{\alpha}_i) + o_{p^*}\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|\right) + o_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|\right) + O_{p^*}\left(\sup_i(\alpha_i^* - \hat{\alpha}_i)^2\right) + O_p\left(\sup_i(\hat{\alpha}_i - \alpha_{i0})^2\right). \quad \text{(A.24)}$$

Now analogous to the proof Theorem 2, define

$$\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i) = \mathbb{H}_{Ti}^{(\beta)*}(\boldsymbol{\theta}_i) - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i\left(\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i) - \frac{\lambda_T}{T}\operatorname{sgn}(\alpha_i)\right) \quad \text{(A.25)}$$

and note that $\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) = O_{p^*}(\lambda_T/T)$. Then equation (A.23) can be rewritten as

$$\left(\bar{\boldsymbol{J}}_i - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i\bar{\boldsymbol{E}}_i'\right)(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) + o_{p^*}\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|\right) + o_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|\right) = \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) + \left(\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right) + O_{p^*}(T^{-1}\lambda_T) + O_{p^*}\left(\sup_i(\alpha_i^* - \hat{\alpha}_i)^2\right) + O_p\left(\sup_i(\hat{\alpha}_i - \alpha_{i0})^2\right). \quad \text{(A.26)}$$

Rearrange and average over i to find

$$\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}} + o_{p^*}\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|\right) + o_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|\right) = \bar{\boldsymbol{D}}_N^{-1}\frac{1}{N}\sum_{i=1}^N \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) + \bar{\boldsymbol{D}}_N^{-1}\frac{1}{N}\sum_{i=1}^N\left(\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) + \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right) + O_{p^*}(T^{-1}\lambda_T) + O_{p^*}\left(\sup_i(\alpha_i^* - \hat{\alpha}_i)^2\right) + O_p\left(\sup_i(\hat{\alpha}_i - \alpha_{i0})^2\right). \quad \text{(A.27)}$$

Next we find the stochastic order of the second term on the right-hand side of (A.27). With $\boldsymbol{X}_{it} = (\boldsymbol{x}_{it}', 1)'$, let $\boldsymbol{X}_{it}'\boldsymbol{\Delta} = \boldsymbol{x}_{it}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (\hat{\alpha}_i - \alpha_{i0})$ and $\boldsymbol{X}_{it}'\boldsymbol{\delta} = \boldsymbol{x}_{it}'(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) + (\alpha_i^* - \hat{\alpha}_i)$ and write $\hat{u}_{it} = u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}$. Define the functions $g_{\boldsymbol{\delta}}(w, u, \boldsymbol{X}, \boldsymbol{\Delta}) = I(w|u + \boldsymbol{X}'\boldsymbol{\Delta}| - \boldsymbol{X}'\boldsymbol{\delta} < 0) - I(w|u + \boldsymbol{X}'\boldsymbol{\Delta}| < 0)$. The class of functions $g_{\boldsymbol{\delta}} - \mathrm{E}^*[g_{\boldsymbol{\delta}}]$ is a bounded, mean-zero VC-subgraph class of functions. Finally, letting $\underline{c} = \min\{c_1, c_2\}$, where c_1, c_2 were used in A2, the unconditional second moment of g_δ satisfies

$$\begin{aligned}
\mathrm{E}\left[(g_{\boldsymbol{\delta}}(\boldsymbol{Z}_{it}))^2\right] &= \mathrm{E}\left[I(|w_{it}||u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}| < |\boldsymbol{X}_{it}'\boldsymbol{\delta}|)\right] \\
&\leq \mathrm{E}\left[I(|u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}| < |\boldsymbol{X}_{it}'\boldsymbol{\delta}|/\underline{c})\right] \\
&= \mathrm{E}\left[F_i(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} + |\boldsymbol{X}_{it}'\boldsymbol{\delta}|/\underline{c}|\boldsymbol{X}_{it}) - F_i(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} - |\boldsymbol{X}_{it}'\boldsymbol{\delta}|/\underline{c}|\boldsymbol{X}_{it})\right] \\
&\leq K(M+1)\|\boldsymbol{\delta}\|, \qquad \text{(A.28)}
\end{aligned}$$

the last inequality holding due to Assumption B4. This implies that $\mathrm{E}^*\left[(g_{\boldsymbol{\delta}}(\boldsymbol{Z}_{it}) - \mathrm{E}^*\left[g_{\boldsymbol{\delta}}(\boldsymbol{Z}_{it})\right])^2\right] \leq K(M+1)\|\boldsymbol{\delta}\|$ with probability approaching 1. Then Proposition B.1 of Kato, Galvao, and Montes-Rojas (2012) implies that with $\delta_N^* = \sup_i |\alpha_i^* - \hat{\alpha}_i| + \|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|$ and $d_{NT}^* = |\log \delta_N^*|/T \vee \sqrt{\delta_N^* |\log \delta_N^*|/T}$,

$$\bar{\boldsymbol{D}}_N^{-1} \frac{1}{N} \sum_{i=1}^{N} \left(\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) + \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right) = O_{p^*}(d_{NT}^*) = o_{p^*}(T^{-1/2}), \quad \text{(A.29)}$$

where the last equality comes from the consistency of $\boldsymbol{\theta}^*$.

Combine (A.27), (A.29), the fact that the first term on the right-hand side of (A.27) is $O_{p^*}((NT)^{-1/2}) = o_{p^*}(T^{-1/2})$ and $\sup_i |\hat{\alpha}_i - \alpha_{i0}| = O_p\left(T^{-1/2}(\log N)^{1/2}\right) = o_p(T^{-1/2})$ to write

$$\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\| = O_{p^*}\left(\sup_i (\alpha_i^* - \hat{\alpha}_i)^2\right) + O_{p^*}(T^{-1}\lambda_T) + o_{p^*}\left(T^{-1/2}\right) + o_p\left(T^{-1/2}\right). \quad \text{(A.30)}$$

Then the preliminary rates of convergence of the coordinates of $\boldsymbol{\theta}_i^*$ can be established similarly to the proof of asymptotic normality of $\hat{\boldsymbol{\theta}}_i$. For example, using (A.22),

$$\sup_i |\alpha_i^* - \hat{\alpha}_i| \leq K \Bigg\{ \sup_i \left| \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T} \operatorname{sgn}(\hat{\alpha}_i) \right| + \sup_i \left| \mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i) - \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i^*) + \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i)\right] \right| \Bigg\} + O_{p^*}(T^{-1}\lambda_T) + o_{p^*}(T^{-1/2}) + o_p(T^{-1/2}) \quad \text{(A.31)}$$

with probability approaching 1. These terms can be bounded by following the calculations similar to the asymptotic normality proof, conditional on the data, using the functions $g_{\boldsymbol{\delta}}(\boldsymbol{Z})$ defined earlier, resulting in $\sup_i |\alpha_i^* - \hat{\alpha}_i| = O_{p^*}(T^{-1/2}(\log N)^{1/2})$. Using (A.30), this implies $\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\| = o_{p^*}(T^{-1/2}(\log N)^{1/2})$. The rest of the proof proceeds as in the proof of asymptotic normality of $\hat{\boldsymbol{\theta}}$, with the addition of the moment conditions on the w_{it} and the convergence of $\bar{\varphi}_i$, $\bar{\boldsymbol{E}}_i$ and $\bar{\boldsymbol{J}}_i$ to their population counterparts for all i using the law of large numbers as $N, T \to \infty$. □

Lemma 2. *Suppose that Assumptions A1-A3 and B1-B3 hold. If $\boldsymbol{\theta}_i$ for $1 \leq i \leq N$ lie in a compact set and $\sup_{N,T} \mathrm{E}\left[|\sqrt{N}\lambda_T/\sqrt{T}|^q\right] < \infty$ for $q > 2$, then $\sup_{N,T} \mathrm{E}^*\left[\|\sqrt{NT}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})\|^q\right] < \infty$.*

Proof of Lemma 2. Follow the steps in the expansions used in Theorem 3 but write out the remainder terms explicitly. Specifically, rewrite (A.21) as

$$\mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i)\right] = -\bar{\boldsymbol{E}}_i'(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) - \bar{\varphi}_i(\alpha_i^* - \hat{\alpha}_i) + (\lambda_T/T)\left(\operatorname{sgn}(\alpha_i^*) - \operatorname{sgn}(\hat{\alpha}_i)\right) + R_i^{(\alpha)*}$$

where

$$R_i^{(\alpha)*} := \mathrm{E}^*\left[\frac{1}{T}\sum_{t=1}^{T} \psi_\tau(w_{it}|\hat{u}_{it}| - \boldsymbol{x}_{it}'(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) - (\alpha_i^* - \hat{\alpha}_i)) - \frac{1}{T}\sum_{t=1}^{T}\psi_\tau(w_{it}|\hat{u}_{it}|)\right] + \bar{\boldsymbol{E}}_i'(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) + \bar{\varphi}_i(\alpha_i^* - \hat{\alpha}_i).$$

Similarly define

$$R_i^{(\beta)*} := \mathrm{E}^* \left[\mathbb{H}_{Ti}^{(\beta)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\beta)*}(\hat{\boldsymbol{\theta}}_i)\right] + \bar{\boldsymbol{J}}_i(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) + \bar{\boldsymbol{E}}_i(\alpha_i^* - \hat{\alpha}_i)$$

which was represented by error terms in equation (A.24) in the proof of Theorem 3. Then (A.27) can be equivalently written

$$\bar{\boldsymbol{D}}_N(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) - \frac{1}{N}\sum_{i=1}^N \left(R_i^{(\beta)*} - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i R_i^{(\alpha)*}\right) + O_{p^*}(\lambda_T/T) = \frac{1}{N}\sum_{i=1}^N \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)$$
$$+ \frac{1}{N}\sum_{i=1}^N \left(\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right). \quad \text{(A.32)}$$

The left-hand side includes the remainder terms, which are functions of the difference between bootstrap parameter estimate and original-sample parameter estimate. Assuming the parameters lie in a compact set implies that the remainder terms are uniformly bounded and have q-th moment. The q-th moment of the other remainder is finite by assumption. The rest of the proof shows that the right hand side is uniformly q-integrable.

Consider the first term on the right-hand side of (A.32), scaled by $\sqrt{NT}$:

$$\sqrt{NT}\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T(\boldsymbol{x}_{it} - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)(\tau - I(w_{it}|\hat{u}_{it}| \le 0)). \quad \text{(A.33)}$$

The bounds on $\boldsymbol{x}_{it}$ and the density of the errors imply that $\bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i$ are bounded. Given the conditions on G_W, the expected value of each summand is zero conditional on the data.

Let $\boldsymbol{X}_{it} = (\boldsymbol{x}_{it}', 1)'$, $\boldsymbol{Z}_{it} = (w_{it}, u_{it}, \boldsymbol{X}_{it})$ and $\boldsymbol{\Delta} \in \mathbb{R}^{p+1}$. Define the class of functions $\mathcal{H} = \{h_{\boldsymbol{\Delta}}(\boldsymbol{Z}) := I(w|u + \boldsymbol{X}'\boldsymbol{\Delta}| < 0) : \boldsymbol{\Delta} \in \mathbb{R}^{p+1}\}$. This class of indicators is a VC subgraph class. To see this, first rewrite

$$\{w|u + \boldsymbol{X}'\boldsymbol{\Delta}| < 0\} = \{w|u + \boldsymbol{X}'\boldsymbol{\Delta}| < 0\} \cap \{w < 0\} \cup \{w|u + \boldsymbol{X}'\boldsymbol{\Delta}| < 0\} \cap \{w > 0\}.$$

For w positive (the opposite case is analogous), the class of sets $\{w|u + \boldsymbol{X}'\boldsymbol{\Delta}| < 0\} \cap \{w > 0\}$ is equivalent to the class $\{|u + \boldsymbol{X}'\boldsymbol{\Delta}| < 0\} = \{u + \boldsymbol{X}'\boldsymbol{\Delta} < 0\} \cap \{-u - \boldsymbol{X}'\boldsymbol{\Delta} < 0\}$. Each of these sets forms a VC class (van der Vaart and Wellner, 1996, Problem 2.6.14) and the class of their intersections is also a VC class (van der Vaart and Wellner, 1996, Lemma 2.6.17). Then the class of unions of sets formed in this way is also a VC class, and $\mathcal{H}$ is a VC subgraph class.

Because of the fact that the indicators in equation (A.33) are a VC subgraph class bounded by 1, their uniform covering number satisfies $\sup_Q N(\epsilon, \mathcal{H}, L_2(Q)) \le A\left(\frac{1}{\epsilon}\right)^v$ for some A, v and $0 < \epsilon < 1$ and Q a probability measure. This implies that

$$J(1, \mathcal{H}) := \sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon\|F\|_{Q,2}, \mathcal{H}, L_2(Q))}\mathrm{d}\epsilon < \infty \quad \text{(A.34)}$$

where the supremum is taken over all discrete probability measures Q (van der Vaart and Wellner, 1996, p. 239). Then Theorem 2.14.1 of van der Vaart and Wellner (1996) implies that there exists a constant C such that

$$\max_{1\le i\le N}\sup_{T\ge 1}\mathrm{E}\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)(\tau-I(w_{it}|\hat{u}_{it}|\le 0))\right\|^q\right]\le CJ(1,\mathcal{H}).$$

Then van der Vaart and Wellner (1996, Problem 2.3.4) implies that

$$\sup_{N,T}\mathrm{E}\left[\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)(\tau-I(w_{it}|\hat{u}_{it}|\le 0))\right\|^q\right]<\infty.$$

For the second term, it is sufficient to consider, for any i,

$$\sup_{T\ge 1}\mathrm{E}\Bigg[\Bigg|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\tau\left(w_{it}|\hat{u}_{it}|-\boldsymbol{x}_{it}'(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}})-(\alpha_i^*-\hat{\alpha}_i)\right)-\psi_\tau\left(w_{it}|\hat{u}_{it}|\right)$$
$$-\mathrm{E}^*\left[\psi_\tau\left(w_{it}|\hat{u}_{it}|-\boldsymbol{x}_{it}'(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}})-(\alpha_i^*-\hat{\alpha}_i)\right)-\psi_\tau\left(w_{it}|\hat{u}_{it}|\right)\right]\Bigg|^q\Bigg].$$

Let $\boldsymbol{X}_{it}=(\boldsymbol{x}_{it}',1)'$, $\boldsymbol{Z}_{it}=(w_{it},u_{it},\boldsymbol{X}_{it})$ and $\boldsymbol{\xi}=(\boldsymbol{\Delta}',\boldsymbol{\delta}')'\in\mathbb{R}^{2(p+1)}$. Define the functions $g_{\boldsymbol{\xi}}(\boldsymbol{Z})=I(w|u+\boldsymbol{X}'\boldsymbol{\Delta}|-\boldsymbol{X}'\boldsymbol{\delta}<0)-I(w|u+\boldsymbol{X}'\boldsymbol{\Delta}|<0)$ and the class of functions $\mathcal{G}=\{g_{\boldsymbol{\xi}}-\mathrm{E}\left[g_{\boldsymbol{\xi}}\right]:\boldsymbol{\xi}\in\mathbb{R}^{2(p+1)}\}$. Then the above display is finite if

$$\mathrm{E}\left[\left|\sup_{\boldsymbol{\xi}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(g_{\boldsymbol{\xi}}(\boldsymbol{Z}_{it})-\mathrm{E}^*\left[g_{\boldsymbol{\xi}}(\boldsymbol{Z}_{it})\right]\right)\right|^q\right]<\infty.$$

However, manipulations similar to the previous step show that $\mathcal{G}$ is also a VC-subgraph class, and therefore, using (A.34) for the class $\mathcal{G}$, we have for another constant C that

$$\max_{1\le i\le N}\sup_{T\ge 1}\mathrm{E}\left[\left\|\sqrt{T}\left(\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i)-\mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i)-\mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i^*)+\mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i)\right]\right)\right\|^q\right]\le CJ(1,\mathcal{G})<\infty.$$

This implies

$$\sup_{N,T}\mathrm{E}\left[\left\|\sqrt{NT}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i)-\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)-\mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*)+\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right)\right\|^q\right]<\infty.$$

The c_r inequality implies that the right-hand side of (A.32) is uniformly q-integrable. Under the assumption that $\bar{\boldsymbol{D}}_N$ is invertible, $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0$ must be as well. □

Proof of Theorem 4. This proof is similar to Theorem 3.2, part (i) of Hagemann (2017). Let $\boldsymbol{Z}_{NT}^*=\sqrt{NT}(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}})$. Theorem 3 shows that $\boldsymbol{Z}_{NT}^*\xrightarrow{d}\boldsymbol{Z}$ in probability, where $\boldsymbol{Z}$ is defined by the condition $\sqrt{NT}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{d}\boldsymbol{Z}$. $\mathrm{E}^*\left[\boldsymbol{Z}_{NT}^*\boldsymbol{Z}_{NT}^{*\prime}\right]\xrightarrow{p}\mathrm{E}\left[\boldsymbol{Z}\boldsymbol{Z}'\right]$ if and only if each coordinate converges in probability, so assume that $p=1$ and we may deal with the 1-dimensional random variables Z_{NT}^* and Z. For any $K>0$, write

$$\left|\mathrm{E}^*\left[Z_{NT}^{*2}\right]-\mathrm{E}\left[Z^2\right]\right|\le\mathrm{E}^*\left[Z_{NT}^{*2}\right]-\mathrm{E}^*\left[\min\{Z_{NT}^{*2},K\}\right]$$
$$+\left|\mathrm{E}^*\left[\min\{Z_{NT}^{*2},K\}\right]-\mathrm{E}\left[\min\{Z^2,K\}\right]\right|+\left|\mathrm{E}\left[\min\{Z^2,K\}\right]-\mathrm{E}\left[Z^2\right]\right|.$$

The portmanteau lemma (van der Vaart, 1998) implies that $\mathrm{E}^*\left[g(Z^*_{NT})\right] \xrightarrow{p} \mathrm{E}\left[g(Z)\right]$ for all continuous and bounded functions g, and that the second term on the right-hand side converges in probability to zero. The first and third terms on the right-hand side are similar; consider just the first term. Note that $\mathrm{E}^*\left[Z^{*2}_{NT}\right] - \mathrm{E}^*\left[\min\{Z^{*2}_{NT}, K\}\right] \leq \mathrm{E}^*\left[Z^{*2}_{NT} I(Z^{*2}_{NT} > K)\right]$. For any $\epsilon > 0$,

$$\mathrm{E}\left[\mathrm{E}^*\left[Z^{*2}_{NT} I(Z^{*2}_{NT} > K)\right]\right] \leq \sup_{N,T} \mathrm{E}\left[Z^{*2(1+\epsilon)}_{NT}\right] K^{-\epsilon}$$

where the expectation on the right-hand side is taken with respect to all the random variables. Lemma 2 (letting $q = 2(1+\epsilon)$ there) implies that the expectation on the right-hand side is finite, so the right-hand side converges to zero as $K \to \infty$. The Markov inequality implies the result. □

Proof of Theorem 5. The proof of this theorem requires minor modifications to that of Theorems 1 and 2. Therefore we only specify the differences here.

To show consistency, first note that Assumption B4 implies Assumption B2, used towards the beginning of the consistency proof. Next, the bound using Hoeffding's inequality must be replaced. Imposing the condition on λ_T and choosing $q = [\sqrt{T}]$ and $s = 2\log N$, Corollary C.1 of Kato, Galvao, and Montes-Rojas (2012) implies that

$$\max_{1\leq i\leq N} \mathrm{P}\left\{ \sup_{\boldsymbol{\theta}_i \in \mathcal{B}_i(\phi)} \left|(\lambda_T/T)(|\alpha_i| - |\alpha_{i0}|)\right| + \left|\Delta_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}\left[\Delta_{Ti}(\boldsymbol{\theta}_i)\right]\right| \geq \epsilon_\phi \right\} = o(N^{-1}),$$

which implies (along with the rest of the argument in Theorem 1) consistency of the estimator.

To show asymptotic normality, there are several terms that should be bounded under the dependent error condition. The proof follows that of Theorem 2 until equation (A.11). The arguments leading to an analog of equation (A.14) are as in the proof of Theorem 5.1 of Kato, Galvao, and Montes-Rojas (2012) — specifically, use Corollary C.1 and Lemma C.1 with $q = [T^c]$ for some sufficiently small $0 < c < 1$ and $s = 2\log N$ to show that

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^N \mathbb{K}^{(\theta)}_{Ti}(\hat{\boldsymbol{\theta}}_i) - K^{(\theta)}_{Ti}(\hat{\boldsymbol{\theta}}_i) - \mathbb{K}^{(\theta)}_{Ti}(\boldsymbol{\theta}_{i0}) + K^{(\theta)}_{Ti}(\boldsymbol{\theta}_{i0})\right\| &= O_p(T^{-1/2}\delta_N^{1/4}(\log N)^{1/2} \vee T^{c-1}\log N) \\ &= o_p(T^{-1/2}(\log N)^{1/2})\end{aligned}$$

and similarly, using the same $q = [T^c]$ and $s = 2\log N$,

$$\begin{aligned}\sup_i \left|\mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\theta}_{i0}) - \frac{\lambda_T}{T}\operatorname{sgn}(\alpha_{i0})\right| &= O_p(T^{-1/2}(\log N)^{1/2}) \\ \sup_i \left\|\mathbb{H}^{(\alpha)}_{Ti}(\hat{\boldsymbol{\alpha}}_i) - H^{(\alpha)}_{Ti}(\hat{\boldsymbol{\alpha}}_i) - \mathbb{H}^{(\alpha)}_{Ti}(\boldsymbol{\alpha}_{i0}) + H^{(\alpha)}_{Ti}(\boldsymbol{\alpha}_{i0})\right\| &= o_p(T^{-1/2}(\log N)^{1/2}).\end{aligned}$$

To show the asymptotic normality of the term analogous to the final sum in the proof of Theorem 2, note that all the $\mathbb{K}^{(\theta)}_{Ti}(\boldsymbol{\theta}_{i0})$ which are defined below equation (A.10) are independent by Assumption C1. For a given i, $(\tau - I(y_{it} \leq \boldsymbol{x}'_{it}\boldsymbol{\beta}_0 + \alpha_{i0}))(\boldsymbol{x}_{it} - \varphi_i^{-1}\boldsymbol{E}_i)$ are uniformly bounded so $\sup_i \mathrm{E}\left[|\mathbb{K}^{(\theta)}_{Ti}(\boldsymbol{\theta}_{i0})|^3\right] < \infty$, while $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \mathrm{Var}(\mathbb{K}^{(\theta)}_{Ti}(\boldsymbol{\theta}_{i0})) = \tilde{\boldsymbol{V}}$ which is positive definite by assumption. □

Proof of Theorem 6. The proof is a modification of the proof of Theorem 3, following the developments in Theorem 5. To save space and avoid repetition, we concentrate our attention on the modifications of the proof.

For consistency we need a bootstrap equivalent of (A.18). Apply the Bernstein inequality for β-mixing sequences in Corollary C.1 of Kato, Galvao, and Montes-Rojas (2012), choosing $q = [\sqrt{T}]$ and $s = 2\log N$. Because of the condition on λ_T we concentrate on the second term of (A.18). Under Assumption C1, we have that

$$\max_{1\leq i\leq N} \mathrm{P}^* \left\{ \sup_{\boldsymbol{\theta}_i, \in \mathcal{B}_i(\phi)} \left| \Delta^{\circ}_{Ti}(\boldsymbol{\theta}_i) - \mathrm{E}^* \left[\Delta^{\circ}_{Ti}(\boldsymbol{\theta}_i) \right] \right| \geq \epsilon_\phi \right\} = o_p \left(N^{-1} \right).$$

The proof of asymptotic normality is analogous to that of Theorem 3 through expansion (A.27). The condition on λ_T and Theorem 2 imply that several of the remainder terms are small, and we need only make one order estimate in (A.29) and two estimates in (A.31) under the β-mixing assumption.

First, find an expression similar to (A.29) under Assumption C1. The calculations in Theorem 3 leading up to (A.28) imply that $\mathrm{E}^* \left[(g_{\boldsymbol{\delta}}(\boldsymbol{Z}_{it}))^2 \right] \leq C\|\boldsymbol{\delta}\|$ with probability approaching 1, and the Cauchy-Schwarz inequality implies similarly that for any $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2$, $\mathrm{E}^* \left[|g_{\boldsymbol{\delta}_1}(\boldsymbol{Z}_{it}) \cdot g_{\boldsymbol{\delta}_2}(\boldsymbol{Z}_{it})|^2 \right] \leq C\|\boldsymbol{\delta}_1\|\|\boldsymbol{\delta}_2\|$ with probability approaching 1. Therefore Lemma C.1 of Kato, Galvao, and Montes-Rojas (2012) implies that for any positive integer q, with $\delta^*_N = \sup_i |\alpha^*_i - \hat{\alpha}_i| + \|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|$,

$$\mathrm{Var}^* \left(\frac{1}{\sqrt{q}} \sum_{t=1}^{q} g_{\boldsymbol{\delta}}(\boldsymbol{Z}_{it}) \right) \leq (\delta^*_N)^{1/2}$$

with probability approaching 1. For some $c \in (0,1)$ let $q = [T^c]$ and $s = 2\log N$, and apply Corollary C.1 of Kato, Galvao, and Montes-Rojas (2012) to find

$$\begin{aligned} &\bar{\boldsymbol{D}}_N^{-1} \frac{1}{N} \sum_{i=1}^{N} \left(\mathbb{K}^{(\theta)*}_{Ti}(\boldsymbol{\theta}^*_i) - \mathbb{K}^{(\theta)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^* \left[\mathbb{K}^{(\theta)*}_{Ti}(\boldsymbol{\theta}^*_i) + \mathbb{K}^{(\theta)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) \right] \right) \\ &\qquad = O_{p^*} \left(T^{-1/2} (\delta^*_N)^{1/4} (\log N)^{1/2} \vee T^{c-1} \log N \right). \qquad \text{(A.35)} \end{aligned}$$

Second, consider the expansion (A.31) under Assumption C1. The second term is $o_{p^*}(T^{-1/2})$ using the result from the previous paragraph. The first term is an average of (under the bootstrap measure) mean-zero terms. It can be verified directly that $\mathrm{E}^* \left[\psi^2_\tau(u^*_{it}) \right] = \tau(1-\tau)$ and $\mathrm{E}^* \left[|\psi_\tau(u^*_{it})\psi_\tau(u^*_{is})|^2 \right] \leq \tau^2(1-\tau)^2$. Then Corollary C.1 of Kato, Galvao, and Montes-Rojas (2012) implies that (using $s = 2\log N$ and $q = [T^c]$)

$$\sup_i \left| \mathbb{H}^{(\alpha)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T} \operatorname{sgn}(\hat{\alpha}_i) \right| = O_{p^*} \left(T^{-1/2} (\log N)^{1/2} \vee T^{c-1} \log N \right). \qquad \text{(A.36)}$$

Now using (A.35) and (A.31) along with the rate condition on N and T and the condition on λ_T, we have (recalling definition (A.25))

$$\sqrt{NT}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) = \bar{\boldsymbol{D}}_N^{-1} \sqrt{NT} \frac{1}{N} \sum_{i=1}^{N} \mathbb{K}^{(\theta)*}_{Ti}(\hat{\boldsymbol{\theta}}_i) + o_{p^*}(1) \qquad \text{(A.37)}$$

Under conditions A1-A3, it is clear that $\mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right] = 0$ for all i, and

$$\mathrm{Var}^*\left(\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right) = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)(\boldsymbol{x}_{it} - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)'\mathrm{E}^*\left[\psi_\tau^2(u_{it}^*)\right]$$
$$+ 2\sum_{j=1}^{T-1}(1 - j/T)(\boldsymbol{x}_{it} - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)(\boldsymbol{x}_{it+j} - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)'\mathrm{E}^*\left[\psi_\tau(u_{it}^*)\psi_\tau(u_{it+j}^*)\right].$$

It can be calculated directly that the expected values in the first sum on the right-hand side are all $\tau(1-\tau)$. Therefore (given the convergence in probability of $\bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i$ to $\varphi_i^{-1}\boldsymbol{E}_i$) for consistent variance estimation it is sufficient to show that,

$$\operatorname*{plim}_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T-1}(1-j/T)\left(\mathrm{E}^*\left[\psi_\tau(u_{it}^*)\psi_\tau(u_{it+j}^*)\right] - \mathrm{E}\left[\psi_\tau(u_{it})\psi_\tau(u_{it+j})|\boldsymbol{x}_{it},\boldsymbol{x}_{it+j}\right]\right) = 0. \tag{A.38}$$

For any (i,t) we have

$$\mathrm{E}^*\left[\psi_\tau(u_{it}^*)\psi_\tau(u_{it+j}^*)\right] = \mathrm{E}^*\left[(\tau - I(u_{it}^* < 0))(\tau - I(u_{it+j}^* < 0))\right]$$
$$= \mathrm{E}^*\left[(\tau - I(w_{it} < 0))(\tau - I(w_{it+j} < 0))\right] = \tau - 2\tau^2 + \mathrm{P}^*\{I(w_{it} < 0, w_{it+j} < 0)\}.$$

Similarly,

$$\mathrm{E}\left[\psi_\tau(u_{it})\psi_\tau(u_{it+j})|\boldsymbol{x}_{it},\boldsymbol{x}_{it+j}\right] = \mathrm{E}\left[(\tau - I(u_{it} < 0))(\tau - I(u_{it+j} < 0))|\boldsymbol{x}_{it},\boldsymbol{x}_{it+j}\right]$$
$$= \tau - 2\tau^2 + \mathrm{P}\left\{u_{it} < 0, u_{it+j} < 0|\boldsymbol{x}_{it}\boldsymbol{x}_{it+j}\right\}.$$

Inserting these expressions in (A.38), it can be seen that A4 implies the variance is correctly estimated.

Finally, we apply a CLT for dependent sequences to (A.37). As in Theorem 5, we check a Lyapunov condition on the sum of the $\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)$ terms by C1. By Assumptions B3, B4, and B5, $\psi_\tau(u_{it}^*)(\boldsymbol{x}_{it} - \bar{\varphi}_i^{-1}\bar{\boldsymbol{E}}_i)$ is uniformly bounded. Moreover, under conditions C1 and C2 and the conditions on w_{it}, $\sup_i \mathrm{E}^*\left[|\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)|^3\right] = O_p(1)$ and $\sum_{i=1}^N \mathrm{E}^*\left[|\mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)|^3\right] = o_p(N^{3/2})$. This implies the result. □

Proof of Theorem 7. The proof of this result is identical to the proof of Theorem 2 through equation (A.11). Lemma S.3 in the supplementary appendix shows that

$$\sup_i |\hat{\alpha}_i - \alpha_{i0}| = O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + T^{-1/2}(\log T)^{1/2} + T^{-1}\lambda_T\right).$$

Rewriting (A.11) using this result (and given that $T^{-2}\lambda_T = o(T^{-1}\lambda_T)$),

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{D}_N^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0}) + \boldsymbol{D}_N^{-1}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{K}_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - K_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0}) + K_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})\right)$$
$$+ O_p(T^{-1}\lambda_T) + O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + T^{-1}\log T\right). \tag{A.39}$$

To show the asymptotic normality of the first term, note that all the $\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})$ are independent across i. For a given i, $(\tau - I(y_{it} \le \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 + \alpha_{i0}))(\boldsymbol{x}_{it} - \varphi_i^{-1}\boldsymbol{E}_i)$ are uniformly bounded so $\sup_i \mathrm{E}\left[|\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})|^3\right] <$

∞, while $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \mathrm{Var}(\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})) = \boldsymbol{V}$ which is positive definite by assumption. These conditions are sufficient to imply that a central limit theorem can be applied to the first term of (A.39). Therefore this term is $O_p((NT)^{-1/2})$.

Lemma S.4 shows that

$$\boldsymbol{D}_N^{-1}\frac{1}{N}\sum_{i=1}^N \left(\mathbb{K}_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - K_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0}) + K_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})\right)$$
$$= O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^{1/2} T^{-1/2}(\log T)^{1/2} + T^{-1}\log T + T^{-2/3}N^{-1/2} + T^{-1}(\log T)^{1/2}\lambda_T^{1/2}\right). \quad \text{(A.40)}$$

Then

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\left((NT)^{-1/2}\right) + O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^{1/2} T^{-1/2}(\log T)^{1/2}\right) + O_p\left(T^{-1}\log T\right)$$
$$+ O_p(T^{-1}\lambda_T) + O_p\left(T^{-1}(\log T)^{1/2}\lambda_T^{1/2}\right).$$

Using the fact that $0 \le \delta \le a + b\delta^{1/2} \Rightarrow 0 \le \delta \le 4\max\{a, b^2\}$ ("fact 1" from Galvao, Gu, and Volgushev (2020)), we may shorten this to

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\left((NT)^{-1/2}\right) + O_p\left(T^{-1}\log T\right) + O_p(T^{-1}\lambda_T) + O_p\left(T^{-1}(\log T)^{1/2}\lambda_T^{1/2}\right).$$

If $\lambda_T = O_p(\log T)$, the final three remainder terms have the same order. If $NT^{-1}(\log T)^2 \to 0$ then the asymptotically normal term dominates, implying the result. □

Proof of Theorem 8. The proof of this theorem is identical to that of Theorem 3 up to (A.27), reprinted here for convenience with some remainder terms changed using the assumption that $\lambda_T = O_p(\log T)$ and what is known of $\hat{\boldsymbol{\theta}}$ from previous theorems:

$$\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}} + o_{p^*}\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|\right) + o_p\left((NT)^{-1/2}\right) = \bar{\boldsymbol{D}}_N^{-1}\frac{1}{N}\sum_{i=1}^N \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)$$
$$+ \bar{\boldsymbol{D}}_N^{-1}\frac{1}{N}\sum_{i=1}^N \left(\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) + \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right)$$
$$+ O_{p^*}(T^{-1}\lambda_T) + O_{p^*}\left(\sup_i(\alpha_i^* - \hat{\alpha}_i)^2\right) + O_{p^*}\left(T^{-1}\log T\right). \quad \text{(A.41)}$$

The inequalities of Lemma S.1.3 of Chao, Volgushev, and Cheng (2017) do not apply to the functions in this expression because of the bootstrap weights in the functions. However, the results of their subsection S.2.1 (which draw on Koltchinskii (2006) and Massart (2000)) may be used to tailor appropriate concentration inequalities.

Rewriting (A.22) with what is known thus far,

$$\alpha_i^* - \hat{\alpha}_i = O_{p^*}\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|\right) + \bar{\varphi}_i^{-1}\left(\mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\operatorname{sgn}(\hat{\alpha}_i)\right)$$
$$+ \bar{\varphi}_i^{-1}\left(\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i)\right]\right)$$
$$+ O_{p^*}(\lambda_T/T) + O_p\left(T^{-1}\log T\right). \quad \text{(A.42)}$$

Noting that

$$\mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\operatorname{sgn}(\hat{\alpha}_i) = \frac{1}{T}\sum_{t=1}^{T}\psi_\tau(w_{it}|\hat{u}_{it}|),$$

this is a sum of mean-zero functions with variance bounded by $\tau(1-\tau)$ and that are members of a VC-subgraph class as described in Lemma 2. Therefore equations S.2.2 and S.2.3 of Chao, Volgushev, and Cheng (2017) may be combined with the union bound to find that

$$\sup_i\left|\mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i) - \frac{\lambda_T}{T}\operatorname{sgn}(\hat{\alpha}_i)\right| = O_{p^*}\left(T^{-1/2}(\log T)^{1/2}\right).$$

Similarly, the terms in the second line of (A.42) were described as $g_{\boldsymbol{\delta}}$ in the proof of Theorem 3. When T^{-1} is smaller than the maximal variance of the $g_{\boldsymbol{\delta}}$ in this class, that is, when $\|C\boldsymbol{\delta}\| > T^{-1}$, S.2.2 and S.2.3 of Chao, Volgushev, and Cheng (2017) may be used again with the union bound to find that, using the notation in the proof of Theorem 3,

$$\sup_i \frac{1}{T}\sum_{t=1}^{T}\left(g_{\boldsymbol{\delta}}(\boldsymbol{Z}_{it}) - \mathrm{E}^*\left[g_{\boldsymbol{\delta}}(\boldsymbol{Z}_{it})\right]\right) = O_{p^*}\left(\|\boldsymbol{\delta}\|^{1/2}T^{-1/2}(\log T)^{1/2} + T^{-1}\log T\right),$$

which in turn imply that

$$\sup_i\left|\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{H}_{Ti}^{(\alpha)*}(\boldsymbol{\theta}_i^*) - \mathbb{H}_{Ti}^{(\alpha)*}(\hat{\boldsymbol{\theta}}_i)\right]\right|$$
$$= O_{p^*}\left(\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\| + \sup_i|\alpha_i^* - \hat{\alpha}_i|\right)^{1/2}T^{-1/2}(\log T)^{1/2} + T^{-1}\log T\right)$$
$$= o_{p^*}(T^{-1/2}(\log T)^{1/2}).$$

These stochastic orders imply that

$$\sup_i|\alpha_i^* - \hat{\alpha}_i| = O_{p^*}\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\| + T^{-1/2}(\log T)^{1/2} + T^{-1}\lambda_T\right).$$

More lengthy calculations that are analogs to Lemmas S.4 and S.5 in the supplemental appendix (conditional on the observations) imply that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) - \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i) - \mathrm{E}^*\left[\mathbb{K}_{Ti}^{(\theta)*}(\boldsymbol{\theta}_i^*) + \mathbb{K}_{Ti}^{(\theta)*}(\hat{\boldsymbol{\theta}}_i)\right]\right)$$
$$= O_{p^*}\left(\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|T^{-1/2}(\log T)^{1/2} + T^{-1}\log T + T^{-2/3}N^{-1/2} + T^{-1}(\log T)^{1/2}\lambda_T^{1/2}\right).$$

Then the rest of the proof goes as in Theorem 7, implying the result. □

References

Belloni, A., and V. Chernozhukov (2011): "ℓ_1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 39(1), 82–130.

Belloni, A., V. Chernozhukov, C. Hansen, and D. Kozbur (2016): "Inference in High-Dimensional Panel Models With an Application to Gun Control," *Journal of Business & Economic Statistics*, 34(4), 590–605.

Bester, C. A., and C. Hansen (2009): "A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects," *Journal of Business & Economic Statistics*, 27(2), 131–148.

Burfisher, M. E., S. Robinson, and K. Thierfelder (2001): "The Impact of NAFTA on the United States," *Journal of Economic Perspectives*, 15(1), 125–144.

Camponovo, L. (2015): "On the validity of the pairs bootstrap for lasso estimators," *Biometrika*, 102(4), 981–987.

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

Chao, S.-K., S. Volgushev, and G. Cheng (2017): "Quantile Processes for Semi and Nonparametric Regression," *Electronic Journal of Statistics*, 11, 3272–3331.

Chen, X., and D. Pouzo (2009): "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics*, 152(1), 46–60.

——— (2015): "Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models," *Econometrica*, 83(3), 1013–1079.

Chetverikov, D., B. Larsen, and C. Palmer (2016): "IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade," *Econometrica*, 84(2), 809–833.

Feng, X., X. He, and J. Hu (2011): "Wild bootstrap for quantile regression," *Biometrika*, 98(4), 995–999.

Galvao, A., J. Gu, and S. Volgushev (2020): "On the Unbiased Asymptotic Normality of Quantile Regression with Fixed Effects," *Journal of Econometrics*, 218, 178–215.

Galvao, G., and K. Kato (2018): "Quantile Regression Methods for Longitudinal Data," in *Handbook of Quantile Regression*, pp. 363–380. CRC Press.

Gonçalves, S., and H. White (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association*, 100(471), 970–979.

Gu, J., and S. Volgushev (2019): "Panel data quantile regression with grouped fixed effects," *Journal of Econometrics*, 213(1), 68 – 91.

Hagemann, A. (2017): "Cluster-Robust Bootstrap Inference in Quantile Regression Models," *Journal of the American Statistical Association*, 112(517), 446–456.

Hahn, J., and Z. Liao (2021): "Bootstrap Standard Error Estimates and Inference," forthcoming, Econometrica.

Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72(4), 1295–1319.

Hakobyan, S., and J. McLaren (2016): "Looking for Local Labor Market Effects of NAFTA," *The Review of Economics and Statistics*, 98(4), 728–741.

Harding, M., and C. Lamarche (2017): "Penalized Quantile Regression with Semiparametric Correlated Effects: An Application with Heterogeneous Preferences," *Journal of Applied Econometrics*, 32(2), 342–358.

——— (2019): "A panel quantile approach to attrition bias in Big Data: Evidence from a randomized experiment," *Journal of Econometrics*, 211(1), 61 – 82.

He, X. (2018): "Resampling Methods," in *Handbook of Quantile Regression*, pp. 7–17. CRC Press.

Kato, K., A. F. Galvao, and G. Montes-Rojas (2012): "Asymptotics for Panel Quantile Regression Models with Individual Effects," *Journal of Econometrics*, 170, 76–91.

Knight, K. (1998): "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *Annals of Statistics*, 26, 755–770.

Knight, K., and W. Fu (2000): "Asymptotics for Lasso-type estimators," *Annals of Statistics*, 28, 1356–1378.

Kock, A. B. (2013): "Oracle Efficient Variable Selection in Random and Fixed Effects Panel Data Models," *Econometric Theory*, 29(1), 115–152.

——— (2016): "Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models," *Journal of Econometrics*, 195(1), 71 – 85.

Kock, A. B., and H. Tang (2019): "Uniform Inference in High-Dimensional Dynamic Panel Data Models with Aproximately Sparse Fixed Effects," *Econometric Theory*, 35(2), 295–359.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

——— (2021): "Quantreg: Quantile Regression," R package version 5.85, `www.r-project.org`.

Koltchinskii, V. (2006): "Local Rademacher complexities and oracle inequalities in risk minimization," *The Annals of Statistics*, 34, 2593–2656.

Lamarche, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408.

Lee, E. R., H. Noh, and B. U. Park (2014): "Model Selection via Bayesian Information Criterion for Quantile Regression Models," *Journal of the American Statistical Association*, 109(505), 216–229.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2018): "Oracle Estimation of a Change Point in High-Dimensional Quantile Regression," *Journal of the American Statistical Association*, 113(523), 1184–1194.

Massart, P. (2000): "About the Constants in Talagrand's Concentration Inequalities for Empirical Processes," *The Annals of Probability*, 28, 863–884.

Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

Su, L., and G. Ju (2018): "Identifying latent grouped patterns in panel data models with interactive fixed effects," *Journal of Econometrics*, 206(2), 554 – 573.

Su, L., Z. Shi, and P. C. B. Phillips (2016): "Identifying Latent Structures in Panel Data," *Econometrica*, 84(6), 2215–2264.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Wang, L. (2013): "The L_1 penalized LAD estimator for high dimensional linear regression," *Journal of Multivariate Analysis*, 120, 135 – 151.

Wang, L. (2019): "L_1-regularized Quantile Regression with Many Regressors under Lean Assumptions," mimeo, University of Minnesota.

Wang, L., I. Van Keilegom, and A. Maidman (2018): "Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors," *Biometrika*, 105(4), 859–872.