SUPPLEMENTARY APPENDIX TO "WILD BOOTSTRAP INFERENCE FOR PENALIZED QUANTILE REGRESSION FOR LONGITUDINAL DATA"*

CARLOS LAMARCHE AND THOMAS PARKER†

S.1. Additional Theoretical Results

Lemma S.1 below implies a natural upper bound for λ_T. If we consider the α_i as parameters associated with indicator functions for individual i in the design matrix, then the column associated with each i has L_1 norm equal to T. In the text we set $\lambda_U = \max\{\tau, 1-\tau\}T$, because otherwise all the individual effects would be set to zero.

Lemma S.1. *Subdivide the covariates for the i-th observation as* $(\boldsymbol{X}_i', x_{pi})' \in \mathbb{R}^p$. *Suppose that the conformable vector of estimates* $(\hat{\boldsymbol{a}}, \hat{b})$ *is defined by*

$$(\hat{\boldsymbol{a}}, \hat{b}) = \underset{\boldsymbol{a}, b \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{X}_i'\boldsymbol{a} - b x_{pi}) + \lambda(\|\boldsymbol{a}\|_1 + |b|). \tag{S.1}$$

Then letting $\boldsymbol{x}_p$ *denote the p-th column of the design matrix,*

$$\max\{\tau, 1-\tau\}\|\boldsymbol{x}_p\|_1 < \lambda \quad \Rightarrow \quad \hat{b} = 0.$$

Proof of Lemma S.1. Note that if

$$\min_{\boldsymbol{a}, b}\left(\sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{X}_i'\boldsymbol{a} - b x_{pi}) + \lambda(\|\boldsymbol{a}\|_1 + |b|)\right) - \min_{\boldsymbol{a}}\left(\sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{X}_i'\boldsymbol{a}) + \lambda\|\boldsymbol{a}\|_1\right) > 0,$$

then it is optimal to set $\hat{b} = 0$. Note that (using the definition of the full solution $(\hat{\boldsymbol{a}}, \hat{b})$)

$$\begin{aligned}
\min_{\boldsymbol{a}, b}\left(\sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{X}_i'\boldsymbol{a} - b x_{pi}) + \lambda(\|\boldsymbol{a}\|_1 + |b|)\right) - \min_{\boldsymbol{a}}\left(\sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{X}_i'\boldsymbol{a}) + \lambda\|\boldsymbol{a}\|_1\right) \geq \\
\sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}} - \hat{b} x_{pi}) + \lambda(\|\hat{\boldsymbol{a}}\|_1 + |\hat{b}|) - \sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}}) - \lambda\|\hat{\boldsymbol{a}}\|_1 \\
= \sum_{i=1}^{N}\Big(\rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}} - \hat{b} x_{pi}) - \rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}})\Big) + \lambda|\hat{b}|.
\end{aligned}$$

Therefore if

$$\sum_{i=1}^{N}\Big(\rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}}) - \rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}} - \hat{b} x_{pi})\Big) < \lambda|\hat{b}|,$$

*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

then $\hat{b} \neq 0$ is not optimal. Applying Lemma S.7 to the left-hand side of the above expression, we have

$$\sum_{i=1}^{N} \Big(\rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}}) - \rho_\tau(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{a}} - \hat{b}x_{pi}) \Big) \leq \max\{\tau, 1-\tau\} \sum_{i=1}^{N} |\hat{b}x_{pi}|$$
$$\leq \max\{\tau, 1-\tau\}|\hat{b}|\|\boldsymbol{x}_p\|_1.$$

Therefore for any $b \neq 0$, the condition

$$\max\{\tau, 1-\tau\}|b|\|\boldsymbol{x}_p\|_1 < \lambda|b| \quad \Leftrightarrow \quad \max\{\tau, 1-\tau\}\|\boldsymbol{x}_p\|_1 < \lambda$$

implies that that b is not an optimizer of the objective function. □

The following lemma collects together two results on expansions that are related to the wild bootstrap method described in the main text.

Lemma S.2. *Let u_{it} have conditional distribution F_i and density functions f_i as described in Assumptions B2 and B4, and suppose that Assumption B3 is satisfied. Let $w_{it} \sim G_W$ be independent of $(u_{it}, \boldsymbol{x}_{it})$ and suppose its distribution satisfies Assumptions A1-A3. Then letting $\boldsymbol{X}_{it} = (\boldsymbol{x}_{it}', 1)'$, under either Assumption B1 or C1:*

(1) *For each i,*

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}^* \left[\int_0^{\boldsymbol{X}_{it}'\boldsymbol{\Delta}} (\psi_\tau(w_{it}|u_{it}| - s) - \psi_\tau(w_{it}|u_{it}|))\mathrm{d}s \right] = -f_i(0)\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\Delta}'\boldsymbol{X}_{it}\boldsymbol{X}_{it}'\boldsymbol{\Delta} + o_p(\|\boldsymbol{\Delta}\|^2).$$

(2) *For each i,*

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}^* \left[\psi_\tau(w_{it}|u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}| - \boldsymbol{X}_{it}'\boldsymbol{\delta}) - \psi_\tau(w_{it}|u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}|) \right]$$
$$= -f_i(0)\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_{it}'\boldsymbol{\delta} + O_p((\|\boldsymbol{\Delta}\| + \|\boldsymbol{\delta}\|)^2).$$

Proof. Both parts of this proof use the identity

$$\psi_\tau(u - s) - \psi_\tau(u) = I(s < u < 0)I(s < 0) - I(0 < u < s)I(s \geq 0). \tag{S.2}$$

First we show part 1. Use (S.2) to write

$$\psi_\tau(w_{it}|u_{it}| - s) - \psi_\tau(w_{it}|u_{it}|) = I(s < w_{it}|u_{it}| < 0)I(s < 0) - I(0 < w_{it}|u_{it}| < s)I(s \geq 0).$$

Then rewrite

$$\mathrm{E}^* \left[\int_0^{\boldsymbol{X}_{it}'\boldsymbol{\Delta}} \psi_\tau(w_{it}|u_{it}| - s) - \psi_\tau(w_{it}|u_{it}|)\mathrm{d}s \right] =$$
$$\mathrm{E}^* \left[\int_{\boldsymbol{X}_{it}'\boldsymbol{\Delta}}^{0} I(s < w_{it}|u_{it}| < 0)\mathrm{d}s \right] I(\boldsymbol{X}_{it}'\boldsymbol{\Delta} < 0) - \mathrm{E}^* \left[\int_0^{\boldsymbol{X}_{it}'\boldsymbol{\Delta}} I(0 < w_{it}|u_{it}| < s)\mathrm{d}s \right] I(\boldsymbol{X}_{it}'\boldsymbol{\Delta} > 0). \tag{S.3}$$

Now focusing on just the first expectation,

$$\mathrm{E}\left[\mathrm{E}^*\left[\int_{\boldsymbol{X}_{it}'\boldsymbol{\Delta}}^0 I(s < w_{it}|u_{it}| < 0)\mathrm{d}s\right]\middle|\boldsymbol{X}_{it}\right] = \int_{-\infty}^0 \int_{\boldsymbol{X}_{it}'\boldsymbol{\Delta}}^0 (F_i(-s/w) - F_i(s/w))\mathrm{d}s\mathrm{d}G_W(w)$$
$$= \int_{-\infty}^0 \int_{\boldsymbol{X}_{it}'\boldsymbol{\Delta}}^0 (f_i(\bar{u}) + f_i(\tilde{u}))(s/w)\mathrm{d}s\mathrm{d}G_W(w)$$

where $\bar{u}$ is between $\boldsymbol{X}_{it}'\boldsymbol{\Delta}$ and 0 and $\tilde{u}$ is between $-\boldsymbol{X}_{it}'\boldsymbol{\Delta}$ and 0. Using Fubini's theorem and the properties of the distribution of w_{it},

$$\mathrm{E}\left[\mathrm{E}^*\left[\int_{\boldsymbol{X}_{it}'\boldsymbol{\Delta}}^0 I(s < w_{it}|u_{it}| < 0)\mathrm{d}s\right]\middle|\boldsymbol{X}_{it}\right] I(\boldsymbol{X}_{it}'\boldsymbol{\Delta} < 0) = -\int_{\boldsymbol{X}_{it}'\boldsymbol{\Delta}}^0 \frac{f(\bar{u}) + f_i(\tilde{u})}{2} s\mathrm{d}s I(\boldsymbol{X}_{it}'\boldsymbol{\Delta} < 0)$$
$$= -\left(f_i(0) + O(|\boldsymbol{X}_{it}'\boldsymbol{\Delta}|)\right)\boldsymbol{\Delta}'\boldsymbol{X}_{it}\boldsymbol{X}_{it}'\boldsymbol{\Delta} I(\boldsymbol{X}_{it}'\boldsymbol{\Delta} < 0).$$

An analogous result holds for the other integral, with $I(\boldsymbol{X}_{it}'\boldsymbol{\Delta} > 0)$. Combining the two results and averaging over t for a given i (under Assumption B3 and either Assumption B1 or C1) implies the first assertion.

To show the next part, again use (S.2) to write

$$\psi_\tau(w_{it}|u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}| - \boldsymbol{X}_{it}'\boldsymbol{\delta}) - \psi_\tau(w_{it}|u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}|)$$
$$= I(\boldsymbol{X}_{it}'\boldsymbol{\delta} < w_{it}|u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}| < 0)I(\boldsymbol{X}_{it}'\boldsymbol{\delta} < 0) - I(0 < w_{it}|u_{it} + \boldsymbol{X}_{it}'\boldsymbol{\Delta}| < \boldsymbol{X}_{it}'\boldsymbol{\delta})I(\boldsymbol{X}_{it}'\boldsymbol{\delta} \geq 0).$$

We have, using Assumption A2,

$$\mathrm{E}\left[\mathrm{E}^*\left[I(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} - \boldsymbol{X}_{it}'\boldsymbol{\delta}/w_{it} < u_{it} < -\boldsymbol{X}_{it}'\boldsymbol{\Delta} + \boldsymbol{X}_{it}'\boldsymbol{\delta}/w_{it})I(w_{it} < 0)\right]|\boldsymbol{X}_{it}\right] I(\boldsymbol{X}_{it}'\boldsymbol{\delta} < 0)$$
$$= \int_{-\infty}^0 \left(F_i(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} + \boldsymbol{X}_{it}'\boldsymbol{\delta}/w) - F_i(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} - \boldsymbol{X}_{it}'\boldsymbol{\delta}/w)\right)\mathrm{d}G_W(w)I(\boldsymbol{X}_{it}'\boldsymbol{\delta} < 0).$$

Expand the terms inside this integral around $(\boldsymbol{\Delta}, \boldsymbol{\delta}) = \mathbf{0}$:

$$F_i(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} + \boldsymbol{X}_{it}'\boldsymbol{\delta}/w) = F_i(0) + f_i(\bar{u})(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} + \boldsymbol{X}_{it}'\boldsymbol{\delta}/w)$$
$$F_i(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} - \boldsymbol{X}_{it}'\boldsymbol{\delta}/w) = F_i(0) + f_i(\tilde{u})(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} - \boldsymbol{X}_{it}'\boldsymbol{\delta}/w),$$

where $\bar{u}$ is between $-\boldsymbol{X}_{it}'\boldsymbol{\Delta} + \boldsymbol{X}_{it}'\boldsymbol{\delta}/w$ and 0 and $\tilde{u}$ is between $-\boldsymbol{X}_{it}'\boldsymbol{\Delta} - \boldsymbol{X}_{it}'\boldsymbol{\delta}/w$ and 0. Using Assumptions B2 and A1-A3,

$$\int_{-\infty}^0 \left(f_i(\bar{u})(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} + \boldsymbol{X}_{it}'\boldsymbol{\delta}/w) - f_i(\tilde{u})(-\boldsymbol{X}_{it}'\boldsymbol{\Delta} - \boldsymbol{X}_{it}'\boldsymbol{\delta}/w)\right)\mathrm{d}G_W(w)I(\boldsymbol{X}_{it}'\boldsymbol{\delta} < 0)$$
$$= \int_{-\infty}^0 \left((f_i(\bar{u}) - f_i(\tilde{u}))(\boldsymbol{X}_{it}'\boldsymbol{\Delta}) - w^{-1}\left(f_i(\bar{u}) + f_i(\tilde{u})\right)(\boldsymbol{X}_{it}'\boldsymbol{\delta})\right)\mathrm{d}G_W(w)I(\boldsymbol{X}_{it}'\boldsymbol{\delta} < 0)$$
$$= \left(-f_i(0)(\boldsymbol{X}_{it}'\boldsymbol{\delta}) + O((|\boldsymbol{X}_{it}'\boldsymbol{\Delta}| + |\boldsymbol{X}_{it}'\boldsymbol{\delta}|)^2)\right) I(\boldsymbol{X}_{it}'\boldsymbol{\delta} < 0). \quad \text{(S.4)}$$

Analogous computations imply

$$\mathrm{E}\left[-\mathrm{E}^*\left[I(0<w_{it}|u_{it}+\boldsymbol{X}_{it}'\boldsymbol{\Delta}|<\boldsymbol{X}_{it}'\boldsymbol{\delta})\right]|\boldsymbol{X}_{it}\right]I(\boldsymbol{X}_{it}'\boldsymbol{\delta}\geq 0)$$
$$=\left(-f_i(0)(\boldsymbol{X}_{it}'\boldsymbol{\delta})+O((|\boldsymbol{X}_{it}'\boldsymbol{\Delta}|+|\boldsymbol{X}_{it}'\boldsymbol{\delta}|)^2)\right)I(\boldsymbol{X}_{it}'\boldsymbol{\delta}\geq 0). \quad \text{(S.5)}$$

Combine equations (S.4) and (S.5), average over t for a given and use Assumption B3 and either of Assumptions B1 or C1 to find the second result. □

For the next lemmas let

$$\|\mathbb{P}_{Ti}-P_i\|_{\mathcal{G}}=\sup_{g\in\mathcal{G}}\left|\frac{1}{T}\sum_{t=1}^{T}\left(g(y_{it},\boldsymbol{X}_{it})-\mathbb{E}[g(y_{it},\boldsymbol{X}_{it})]\right)\right|$$

and as in Galvao, Gu, and Volgushev (2020), define

$$\mathcal{G}_1=\left\{(y,\boldsymbol{X})\mapsto\boldsymbol{a}'\boldsymbol{X}(I(y\leq\boldsymbol{b}'\boldsymbol{X})-\tau)I(\|\boldsymbol{X}\|\leq M):\boldsymbol{b}\in\mathbb{R}^{p+1},\boldsymbol{a}\in\mathcal{S}^{p+1}\right\}, \quad \text{(S.6)}$$

where $\boldsymbol{X}=(\boldsymbol{x}',1)'$, and

$$\mathcal{G}_2(\delta)=\left\{(y,\boldsymbol{X})\mapsto\boldsymbol{a}'\boldsymbol{X}(I(y\leq\boldsymbol{X}'\boldsymbol{b}_1)-I(y\leq\boldsymbol{X}'\boldsymbol{b}_2))I(\|\boldsymbol{X}\|\leq M):\right.$$
$$\left.\boldsymbol{b}_1,\boldsymbol{b}_2\in\mathbb{R}^{p+1},\|\boldsymbol{b}_1-\boldsymbol{b}_2\|\leq\delta,\boldsymbol{a}\in\mathcal{S}^{p+1}\right\}. \quad \text{(S.7)}$$

Some lemmas below rely on an infeasible estimate of α_{i0}. For each i, let

$$\tilde{\alpha}_i=\operatorname*{argmin}_a\sum_{t=1}^{T}\rho_\tau(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{\beta}_0-a)+\lambda_T|a|. \quad \text{(S.8)}$$

The $\{\tilde{\alpha}_i\}$ differ from $\{\hat{\alpha}_i\}$ because the latter are all solutions to optimization problems like (S.8) but with $\hat{\boldsymbol{\beta}}$ in the place of $\boldsymbol{\beta}_0$.

Lemma S.3. *Under Assumptions B1 and B3-B5,*

$$\sup_i|\hat{\alpha}_i-\alpha_{i0}|=O_p\left(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|+T^{-1/2}(\log T)^{1/2}+T^{-1}\lambda_T\right). \quad \text{(S.9)}$$

Proof of Lemma S.3. Equation (A.5) from the proof of Theorem 2 implies (under Assumption B4 and using (A.10))

$$\sup_i|\hat{\alpha}_i-\alpha_{i0}|=O_p\left(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|\right)+O_p\left(\sup_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})-\frac{\lambda_T}{T}\operatorname{sgn}(\alpha_{i0})\right)\right)$$
$$+O_p\left(\sup_i\left(\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i)-H_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i)-\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})+H_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})\right)\right)+O_p\left(T^{-1}\lambda_T\right). \quad \text{(S.10)}$$

Note that the expected value of $\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})-(\lambda_T/T)\operatorname{sgn}(\alpha_{i0})=\frac{1}{T}\sum_t\psi_\tau(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{\beta}_0-\alpha_{i0})$ is zero for all i. Setting (their notation first, ours second) $m=p+1$, $n=T$ and $\xi_m=M+1$, and using $\kappa_n=C\log T$ with $C>1$, Lemma S.1.3 of Chao, Volgushev, and Cheng (2017) and the union bound imply that the right-hand side of (S.10) satisfies

$$\sup_i\left|\mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})-(\lambda_T/T)\operatorname{sgn}(\alpha_{i0})\right|=O_p\left(\sup_i\|\mathbb{P}_{Ti}-P_i\|_{\mathcal{G}_1}\right)=O_p\left(T^{-1/2}(\log T)^{1/2}\right). \quad \text{(S.11)}$$

Next, Lemma S.1.3 from Chao, Volgushev, and Cheng (2017) may be used again (with the same constants) to find

$$\begin{aligned}
&\sup_i \left|\mathbb{H}_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - H_{Ti}^{(\alpha)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{H}_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0}) + H_{Ti}^{(\alpha)}(\boldsymbol{\theta}_{i0})\right| \\
&\qquad = O_p\left(\sup_i \|\mathbb{P}_{Ti} - P_i\|_{\mathcal{G}_2(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|+\sup_i|\hat{\alpha}_i-\alpha_{i0}|)}\right) \\
&\qquad = O_p\left((\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \sup_i|\hat{\alpha}_i - \alpha_{i0}|)^{1/2}T^{-1/2}(\log T)^{1/2} + T^{-1}\log T\right) \\
&\qquad = o_p(T^{-1/2}(\log T)^{1/2}) \qquad \text{(S.12)}
\end{aligned}$$

by the consistency of $\hat{\boldsymbol{\theta}}_i$. Using (S.11) and (S.12) in (S.10) implies the result. □

Lemma S.4. *Under Assumptions B1 and B3-B6,*

$$\begin{aligned}
&\boldsymbol{D}_N^{-1}\frac{1}{N}\sum_{i=1}^N\left(\mathbb{K}_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - K_{Ti}^{(\theta)}(\hat{\boldsymbol{\theta}}_i) - \mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0}) + K_{Ti}^{(\theta)}(\boldsymbol{\theta}_{i0})\right) \\
&\quad = O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^{1/2}T^{-1/2}(\log T)^{1/2} + T^{-1}\log T + T^{-2/3}N^{-1/2} + T^{-1}(\log T)^{1/2}\lambda_T^{1/2}\right). \qquad \text{(S.13)}
\end{aligned}$$

Proof of Lemma S.4. First, for ease of notation define

$$\frac{1}{N}\sum_{i=1}^N\left(\mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_i) - K_{Ti}^{(\theta)}(\boldsymbol{\theta}_i) - \mathbb{K}_{Ti}^{(\theta)}(\boldsymbol{\theta}_i') + K_{Ti}^{(\theta)}(\boldsymbol{\theta}_i')\right) := \frac{1}{N}\sum_{i=1}^N \mathcal{K}_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_i').$$

Given the assumed positive definiteness of $\boldsymbol{D}_N$, we may focus on the stochastic order of this average. Recalling that $\tilde{\alpha}_i$ was defined in (S.8), write

$$\frac{1}{N}\sum_{i=1}^N \mathcal{K}_i(\hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_{i0}) = \frac{1}{N}\sum_{i=1}^N \mathcal{K}_i(\hat{\boldsymbol{\theta}}_i, (\boldsymbol{\beta}_0, \tilde{\alpha}_i)) + \frac{1}{N}\sum_{i=1}^N \mathcal{K}_i((\boldsymbol{\beta}_0, \tilde{\alpha}_i), \boldsymbol{\theta}_{i0}). \qquad \text{(S.14)}$$

Suppose that the assumptions of Theorem 2 are satisfied. Recalling the definition of $\mathcal{G}_2(\delta)$ in (S.7),

$$\begin{aligned}
\sup_i \mathcal{K}_i(\hat{\boldsymbol{\theta}}_i, (\boldsymbol{\beta}_0, \tilde{\alpha}_i)) &= O_p\left(\sup_i \|\mathbb{P}_{Ti} - P_i\|_{\mathcal{G}_2(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|+\sup_i|\hat{\alpha}_i-\tilde{\alpha}_i|)}\right) \\
&= O_p\left((\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \sup_i|\hat{\alpha}_i - \tilde{\alpha}_i|)^{1/2}T^{-1/2}(\log T)^{1/2} + T^{-1}\log T\right),
\end{aligned}$$

where the second estimate is a result of Lemma S.1.3 of Chao, Volgushev, and Cheng (2017) with $m = p+1$, $\xi_m = M$ and $\kappa_n = C\log T$, using the union bound for the supremum. Therefore Lemma S.5 implies that

$$\sup_i \mathcal{K}_i(\hat{\boldsymbol{\theta}}_i, (\boldsymbol{\beta}_0, \tilde{\alpha}_i)) = O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^{1/2}T^{-1/2}(\log T)^{1/2} + T^{-1}\log T + T^{-1}(\log T)^{1/2}\lambda_T^{1/2}\right). \qquad \text{(S.15)}$$

Next we require the stochastic order of $\sup_i \mathcal{K}_i((\boldsymbol{\beta}_0, \tilde{\alpha}_i), \boldsymbol{\theta}_{i0})$. Note that the $\{\mathcal{K}_i((\boldsymbol{\beta}_0, \tilde{\alpha}_i), \boldsymbol{\theta}_{i0})\}_i$ are independent and that

$$\mathcal{K}_i((\boldsymbol{\beta}_0, \tilde{\alpha}_i), \boldsymbol{\theta}_{i0}) = O_p\left(\|\mathbb{P}_{Ti} - P_i\|_{\mathcal{G}_2(|\tilde{\alpha}_i-\alpha_{i0}|)}\right).$$

Lemma 3 of Galvao, Gu, and Volgushev (2020) shows that $\mathrm{E}\left[\frac{1}{N}\sum_i \mathcal{K}_i((\boldsymbol{\beta}_0,\tilde{\alpha}_i),\boldsymbol{\theta}_{i0})\right] = O_p(T^{-1}\log T)$. Consider bounding the order of the variance of this average. By Assumption B3, $\sup_i \|\mathcal{K}_i((\boldsymbol{\beta}_0,\tilde{\alpha}_i),\boldsymbol{\theta}_{i0})\| \leq 4M$. In addition, we have

$$\begin{aligned}
&\mathrm{P}\left\{\sup_i \|\mathcal{K}_i((\boldsymbol{\beta}_0,\tilde{\alpha}_i),\boldsymbol{\theta}_{i0})\| > T^{-2/3}\right\} \\
&\qquad \leq \mathrm{P}\left\{\sup_i |\tilde{\alpha}_i - \alpha_{i0}| > cT^{-1/2}(\log T)^{1/2}\right\} + \mathrm{P}\left\{C\|\mathbb{P}_{Ti} - P_i\|_{\mathcal{G}_2(cT^{-1/2}(\log T)^{1/2})} > T^{-2/3}\right\} \\
&\qquad\qquad = O(T^{-2}).
\end{aligned}$$

The above order estimate uses Lemma S.6 with $\kappa = 2$ for the first term. It uses Lemma S.1.3 of Chao, Volgushev, and Cheng (2017) for the second, setting $\xi_n = M+1$, $m = p+1$, $\kappa_n = 2\log T$ and $\delta_n = cT^{-1/2}(\log T)^{1/2}$ (their notation first, ours second), noting that $T^{-3/4}(\log T)^{3/4} = o(T^{-2/3})$. Then the variance of one term in the average, writing $\mathcal{K}_i = \mathcal{K}_i((\boldsymbol{\beta}_0,\tilde{\alpha}_i),\boldsymbol{\theta}_{i0})$, is bounded by

$$\begin{aligned}
\sup_i \mathrm{Var}(\mathcal{K}_i) &\leq \sup_i \mathrm{E}\left[\mathcal{K}_i^2 I(|\mathcal{K}_i| > T^{-2/3}) + \mathcal{K}_i^2 I(|\mathcal{K}_i| \leq T^{-2/3})\right] \\
&\leq 16M^2 \sup_i \mathrm{P}\left\{|\mathcal{K}_i| > T^{-2/3}\right\} + T^{-4/3} = O(T^{-4/3}).
\end{aligned}$$

Then using independence over i and $\mathrm{E}\left[|X|\right] \leq |\mathrm{E}\left[X\right]| + \sqrt{\mathrm{Var}(X)}$,

$$\frac{1}{N}\sum_{i=1}^N \mathcal{K}_i((\boldsymbol{\beta}_0,\tilde{\alpha}_i),\boldsymbol{\theta}_{i0}) = O_p\left(T^{-1}\log T + T^{-2/3}N^{-1/2}\right). \tag{S.16}$$

Use (S.16) and (S.15) in (S.14) to find the result. □

Lemma S.5. *Recall the definition of $\tilde{\alpha}_i$ from* (S.8)*. Under Assumptions B1 and B3-B5,*

$$\sup_i |\hat{\alpha}_i - \tilde{\alpha}_i| = O_p\left(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + T^{-1}\log T + T^{-1}\lambda_T\right). \tag{S.17}$$

Proof of Lemma S.5. For any value of $\boldsymbol{\beta}$ define the empirical CDF of $\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}\}_t$ for unit i by

$$\hat{\mathbb{F}}_{iT}(y,\boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^T I(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} \leq y).$$

Given any value of $\boldsymbol{\beta}$, the solution to $\min_a \sum_{t=1}^T \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - a) + \lambda_T|a|$ is a penalized sample quantile from $\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}\}_{t=1}^T$: the solution a_i^* satisfies

$$\left|\hat{\mathbb{F}}_{iT}(a_i^*,\boldsymbol{\beta}) - \tau + (\lambda_T/T)\,\mathrm{sgn}(a_i^*)\right| \leq 1/T \ a.s. \tag{S.18}$$

That is, a_i^* lies between the $(\tau-(\lambda_T+1)/T)$-th and $(\tau+(\lambda_T+1)/T)$-th sample quantiles of $\{y_{it}-\boldsymbol{x}_{it}'\boldsymbol{\beta}\}_t$. Therefore

$$\left|\hat{\mathbb{F}}_{iT}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}}) - \hat{\mathbb{F}}_{iT}(\tilde{\alpha}_i,\boldsymbol{\beta}_0)\right| = O_p(T^{-1}\lambda_T).$$

Given this, the rest of the proof follows the same steps as the proof of Lemma 7 in Galvao, Gu, and Volgushev (2020), leading to

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

□

The following lemma about penalized sample quantile estimates is analogous to classical results about sample quantiles as in Serfling (1980).

Lemma S.6. *Suppose that Assumptions B1 and B3-B6 hold. Then there is a constant $c > 0$ not depending on i, N or T such that*

$$\mathrm{P}\left\{|\tilde{\alpha}_i - \alpha_{i0}| > c\kappa^{1/2}T^{-1/2}(\log T)^{1/2}\right\} = O(T^{-\kappa}).$$

Proof of Lemma S.6. As in the proof of Lemma S.5, let $\hat{\mathbb{F}}_{iT}(y, \boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T} I(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} \leq y)$. Furthermore let $F_{iT}(y, \boldsymbol{\beta}) = \mathrm{E}\left[\hat{\mathbb{F}}_{iT}(y, \boldsymbol{\beta})\right]$. Given $\boldsymbol{\beta}_0$, the solution $\tilde{\alpha}_i$ for sufficiently large T (assuming $\lambda_T = o_p(T)$) satisfies

$$\left|\hat{\mathbb{F}}_{iT}(\tilde{\alpha}_i, \boldsymbol{\beta}_0) - \tau + (\lambda_T/T)\,\mathrm{sgn}(\tilde{\alpha}_i)\right| \leq 1/T \ a.s. \tag{S.19}$$

Fix $\epsilon > 0$ and note that $\mathrm{P}\{|\tilde{\alpha}_i - \alpha_{i0}| > \epsilon\} = \mathrm{P}\{\tilde{\alpha}_i > \alpha_{i0} + \epsilon\} + \mathrm{P}\{\tilde{\alpha}_i < \alpha_{i0} - \epsilon\}$. Since (S.19) implies that $\hat{\mathbb{F}}_{iT}(\tilde{\alpha}_i, \boldsymbol{\beta}_0) \leq \tau + (\lambda_T + 1)/T$, we may write

$$\begin{aligned}
\mathrm{P}\{\tilde{\alpha}_i > \alpha_{i0} + \epsilon\} &= \mathrm{P}\left\{\hat{\mathbb{F}}_{iT}(\tilde{\alpha}_i, \boldsymbol{\beta}_0) > \hat{\mathbb{F}}_{iT}(\alpha_{i0} + \epsilon, \boldsymbol{\beta}_0)\right\} \qquad \text{(S.20)}\\
&\leq \mathrm{P}\left\{T\tau + (\lambda_T + 1) > \sum_{t=1}^{T} I(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 \leq \alpha_{i0} + \epsilon)\right\}\\
&= \mathrm{P}\left\{\sum_{t=1}^{T} I(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 > \alpha_{i0} + \epsilon) + (\lambda_T + 1) > T(1 - \tau)\right\}.
\end{aligned}$$

Letting $v_{it} = I(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 > \alpha_{i0} + \epsilon)$, rewrite this as

$$\begin{aligned}
&= \mathrm{P}\left\{\sum_{t=1}^{T}(v_{it} - \mathrm{E}[v_{it}]) + (\lambda_T + 1) > TF_{iT}(\alpha_{i0} + \epsilon, \boldsymbol{\beta}_0) - T\tau\right\}\\
&\leq \mathrm{P}\left\{\sum_{t=1}^{T}(v_{it} - \mathrm{E}[v_{it}]) > TF_{iT}(\alpha_{i0} + \epsilon, \boldsymbol{\beta}_0) - T\tau\right\}\\
&\qquad + \mathrm{P}\{\lambda_T + 1 > TF_{iT}(\alpha_{i0} + \epsilon, \boldsymbol{\beta}_0) - T\tau\}. \qquad \text{(S.21)}
\end{aligned}$$

An analogous argument with $\tilde{v}_{it} = I(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 \leq \alpha_{i0} - \epsilon)$ implies that

$$\begin{aligned}
\mathrm{P}\{\tilde{\alpha}_i < \alpha_{i0} - \epsilon\} \leq \mathrm{P}&\left\{\sum_{t=1}^{T}(\tilde{v}_{it} - \mathrm{E}[\tilde{v}_{it}]) > T\tau - TF_{iT}(\alpha_{i0} - \epsilon, \boldsymbol{\beta}_0)\right\}\\
&+ \mathrm{P}\{\lambda_T + 1 > T\tau - TF_{iT}(\alpha_{i0} - \epsilon, \boldsymbol{\beta}_0)\}. \qquad \text{(S.22)}
\end{aligned}$$

Define $\delta_{iT} = \delta_{iT}(\epsilon)$ by

$$\delta_{iT} = \min\left\{F_{iT}(\alpha_{i0}+\epsilon,\boldsymbol{\beta}_0)-\tau, \tau - F_{iT}(\alpha_{i0}-\epsilon,\boldsymbol{\beta}_0)\right\}.$$

Applying Hoeffding's inequality to both (S.21) and (S.22) implies

$$\mathrm{P}\left\{|\tilde{\alpha}_i-\alpha_{i0}|>\epsilon\right\} \le 2e^{-2T\delta_{iT}^2} + 2\mathrm{P}\left\{\lambda_T+1>T\delta_{iT}\right\}. \tag{S.23}$$

Next, given κ in B6, define $\epsilon_T = \underline{f}^{-1}\kappa^{1/2}T^{-1/2}(\log T)^{1/2}$ and consider bounding $\mathrm{P}\left\{|\tilde{\alpha}_i-\alpha_{i0}|>\epsilon_T\right\}$. Note that $F_{iT}(\alpha_{i0}+u,\boldsymbol{\beta}_0) = \mathrm{E}\left[F_{u_{it}|\boldsymbol{x}_{it}}(u|\boldsymbol{x}_{it})\right]$, and Assumption B4 implies that $\underline{f}>0$ exists. As T grows large, again under Assumption B4, $F_{iT}(\alpha_{i0}+\epsilon_T,\boldsymbol{\beta}_0)-\tau = \mathrm{E}\left[f_i(0|\boldsymbol{x}_{it})\right]\epsilon_T + o(\epsilon_T)$, implying that for given constant c, for all T large enough,

$$F_{iT}(\alpha_{i0}+\epsilon_T,\boldsymbol{\beta}_0)-\tau \ge c\kappa^{1/2}T^{-1/2}(\log T)^{1/2}$$

and similarly

$$\tau - F_{iT}(\alpha_{i0}-\epsilon_T,\boldsymbol{\beta}_0) \ge c\kappa^{1/2}T^{-1/2}(\log T)^{1/2}.$$

Therefore the definition of δ_{iT} using ϵ_T implies that $2e^{-2T\delta_{iT}^2} = O(T^{-\kappa})$. Finally, given c, for large enough T we have

$$\mathrm{P}\left\{\lambda_T+1>T\delta_{iT}(\epsilon_T)\right\} \le \mathrm{P}\left\{\lambda_T+1>c\kappa^{1/2}T^{1/2}(\log T)^{1/2}\right\},$$

and by Assumption B6 we may choose c such that the latter sequence of probabilities is $O(T^{-\kappa})$. □

Remark S.1. Condition B6 is nearly equivalent to making the assumption that λ_T behaves like the sum of independent subgaussian random variables. To see this, suppose that with $\mu_T = \mathrm{E}\left[\lambda_T\right]$ and (given κ) $\sigma_T = \sqrt{T/2\kappa}$, we have the Hoeffding bound $\mathrm{P}\left\{(\lambda_T-\mu_T)\ge t\right\} \le \exp\{-t^2/2\sigma_T^2\}$ for all $t>0$. Then $\mathrm{P}\left\{(\lambda_T-\mu_T)>cT^{1/2}(\log T)^{1/2}\right\} \le T^{-\kappa}$. If, in addition, $\mu_T = o(T^{1/2}(\log T)^{1/2})$, then this implies our assumption.

The following lemma shows that the check function satisfies a triangle inequality, and a sort of reverse triangle inequality. The inequality $|\rho_\tau(u)-\rho_\tau(v)| < |u-v|$ for $\tau\in(0,1)$ is used often in the quantile regression literature, but for the computational property of the penalized estimator described above in Lemma S.1, a sharp inequality is required, which is what is shown in the second part of the following lemma.

Lemma S.7. *Let $\rho_\tau(u) = u(\tau - I(u<0))$ for $\tau\in(0,1)$ and $u\in\mathbb{R}$. Then*

(1) $\rho_\tau(u+v) \le \rho_\tau(u)+\rho_\tau(v)$
(2) $|\rho_\tau(u)-\rho_\tau(v)| \le \max\{\tau,1-\tau\}|u-v|$.

Proof of Lemma S.7. It can be verified that $\rho_\tau(u) = \max\{(\tau-1)u,\tau u\}$. This implies both $(\tau-1)u \le \rho_\tau(u)$ and $\tau u \le \rho_\tau(u)$. Therefore $\tau(u+v) = \tau u+\tau v \le \rho_\tau(u)+\rho_\tau(v)$ and $(\tau-1)(u+v) = (\tau-1)u + (\tau-1)v \le \rho_\tau(u)+\rho_\tau(v)$, which together imply

$$\rho_\tau(u+v) = \max\{(\tau-1)(u+v),\tau(u+v)\} \le \rho_\tau(u)+\rho_\tau(v).$$

Next, this inequality implies $\rho_\tau(u) \leq \rho_\tau(u-v) + \rho_\tau(v)$ and $\rho_\tau(v) \leq \rho_\tau(v-u) + \rho_\tau(u)$. Then

$$\rho_\tau(u) - \rho_\tau(v) \leq \rho_\tau(u-v) = \max\{(\tau-1)(u-v), \tau(u-v)\} \leq \max\{\tau, 1-\tau\}|u-v|$$

and similarly, $\rho_\tau(v) - \rho_\tau(u) \leq \max\{\tau, 1-\tau\}|u-v|$. This implies the result. □

S.2. On the cross-sectional pairs bootstrap with fixed N and T

In this section, we offer a heuristic illustration of some problems with using a cross-sectional pairs bootstrap for the penalized quantile regression estimator.

Fix N and T and assume that all $\alpha_{i0} \neq 0$ for simplicity. The assumption on α_{i0} reflects the fact that we make no sparsity assumptions in our analysis (see Knight and Fu (2000) for analogous expressions with some $\alpha_{i0} = 0$). Define $\boldsymbol{\delta} = \sqrt{NT}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and $\boldsymbol{\eta}$ by $\eta_i = \sqrt{T}(\alpha_i - \alpha_{i0})$ for $i = 1, \ldots N$. Then let

$$\mathbb{V}_T(\boldsymbol{\delta}, \boldsymbol{\eta}) = \sum_{i=1}^N \sum_{t=1}^T \left\{ \rho_\tau\left(u_{it} - \frac{\boldsymbol{x}_{it}'\boldsymbol{\delta}}{\sqrt{NT}} - \frac{\eta_i}{\sqrt{T}}\right) - \rho_\tau(u_{it}) \right\} + \lambda_T \sum_{i=1}^N \left\{ \left|\alpha_{i0} + \frac{\eta_i}{\sqrt{T}}\right| - |\alpha_{i0}| \right\}, \tag{S.24}$$

where $u_{it} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 - \alpha_{i0}$. This objective function is equivalent to (2.2) in the main text. Analysis like that of of Koenker (2004) shows that when T is large, letting $f_i = f_{u_{it}|\boldsymbol{x}_{it}}$ and defining $\boldsymbol{\gamma}_i = (\boldsymbol{\delta}'/\sqrt{N}, \eta_i)'$, and letting $A \approx B$ mean that A is approximately distributed as B,

$$\mathbb{V}_T(\boldsymbol{\delta}, \boldsymbol{\eta}) \approx -\sum_{i=1}^N \boldsymbol{\gamma}_i'\boldsymbol{B}_{Ti} + \frac{1}{2}\sum_{i=1}^N \boldsymbol{\gamma}_i'\boldsymbol{D}_{Ti}\boldsymbol{\gamma}_i + \frac{\lambda_T}{\sqrt{T}}\sum_{i=1}^N \eta_i \operatorname{sgn}(\alpha_{i0}), \tag{S.25}$$

where

$$\boldsymbol{B}_{Ti} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \begin{bmatrix} \boldsymbol{x}_{it} \\ 1 \end{bmatrix} \psi_\tau(u_{it}), \qquad \boldsymbol{D}_{Ti} = \frac{1}{T}\sum_{t=1}^T f_i(0|\boldsymbol{x}_{it}) \begin{bmatrix} \boldsymbol{x}_{it}\boldsymbol{x}_{it}' & \boldsymbol{x}_{it} \\ \boldsymbol{x}_{it}' & 1 \end{bmatrix}.$$

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, a bootstrapped estimate, the minimizer of the bootstrap objective function, solves

$$\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\alpha}}'\right)' = \operatorname*{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i=1}^N n_i^* \sum_{t=1}^T \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i\right) + \lambda_T \sum_{i=1}^N n_i^*|\alpha_i|. \tag{S.26}$$

Recenter (S.26) employing $\hat{\boldsymbol{\theta}}$. We find a bootstrap analog of the original objective function (2.3), denoting $\hat{u}_{it} = y_{it} - \hat{\boldsymbol{\beta}}'\boldsymbol{x}_{it} - \hat{\alpha}_i$:

$$\tilde{\mathbb{V}}_T(\boldsymbol{\delta}, \boldsymbol{\eta}) = \sum_{i=1}^N n_i^* \sum_{t=1}^T \left\{ \rho_\tau\left(\hat{u}_{it} - \frac{\boldsymbol{x}_{it}'\boldsymbol{\delta}}{\sqrt{NT}} - \frac{\eta_i}{\sqrt{T}}\right) - \rho_\tau(\hat{u}_{it}) \right\} + \lambda_T \sum_{i=1}^N n_i^* \left\{ \left|\hat{\alpha}_i + \frac{\eta_i}{\sqrt{T}}\right| - |\hat{\alpha}_i| \right\}. \tag{S.27}$$

Then

$$\tilde{\mathbb{V}}_T(\boldsymbol{\delta}, \boldsymbol{\eta}) \approx -\sum_{i=1}^N n_i^*\boldsymbol{\gamma}_i'\tilde{\boldsymbol{B}}_{Ti} + \frac{1}{2}\sum_{i=1}^N n_i^*\boldsymbol{\gamma}_i'\tilde{\boldsymbol{D}}_{Ti}\boldsymbol{\gamma}_i + \frac{\lambda_T}{\sqrt{T}}\sum_{i=1}^N n_i^*\eta_i \operatorname{sgn}(\alpha_{i0}) \tag{S.28}$$

where

$$\tilde{\boldsymbol{B}}_{Ti} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} \boldsymbol{x}_{it} \\ 1 \end{bmatrix} \psi_\tau(u_{it} - \boldsymbol{x}_{it}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - (\hat{\alpha}_i - \alpha_{i0})),$$

$$\tilde{\boldsymbol{D}}_{Ti} = \frac{1}{T} \sum_{t=1}^{T} f_i(\boldsymbol{x}_{it}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (\hat{\alpha}_i - \alpha_{i0})) \begin{bmatrix} \boldsymbol{x}_{it}\boldsymbol{x}_{it}' & \boldsymbol{x}_{it} \\ \boldsymbol{x}_{it}' & 1 \end{bmatrix}.$$

As in Section 2.2, there are two key differences between expressions (S.25) and (S.28). First, $\tilde{\boldsymbol{B}}_{Ti} \neq \boldsymbol{B}_{Ti}$ and $\tilde{\boldsymbol{D}}_{Ti} \neq \boldsymbol{D}_{Ti}$ due to the fact that recentering uses $\hat{\boldsymbol{\theta}}$, which is biased since the model implies that $\mathrm{E}\left[\psi_\tau(u_{it})\right] = 0$.

Second, there is a problem with variability in the penalty term. 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, let $\mathcal{A} = \{i : n_i^* > 0\}$ denote the "active" set of units that are included in the penalty term in (S.26). In each bootstrap repetition, $\mathrm{card}(\mathcal{A}) < N$, potentially changing the penalty significantly and leading to solutions $\tilde{\boldsymbol{\theta}}$ that are very different than the minimizer $\hat{\boldsymbol{\theta}}$.

S.3. Additional Simulation Results

S.3.1. **Finite Sample Performance of the Penalized Estimator.** Figure S.1 shows the bias and root mean squared error (RMSE) of the penalized and fixed effects estimator for the slope parameter. We use the location-scale shift model considered in Section 4 of Kato, Galvao, and Montes-Rojas (2012). The variables are generated as in their second specification. The parameter of interest is $\beta(\tau) = 1 + 0.5F_u(\tau)^{-1}$, where F_u is the distribution of the error term. The model is estimated at $\tau = 0.75$ considering that the error term, u_{it}, is distributed as $\mathcal{N}(0,1)$ or χ^2_3.

The panels in Figure S.1 show that the fixed effects quantile regression (FEQR) estimator is biased when $N = 100$ and $T = 5$. The extent of the bias varies with the distribution of the error term. Note in particular that the bias of the fixed effects estimator is -0.28 (or 9%) when $u_{it} \sim \chi^2_3$, which is consistent with the results in Table 4 in Kato, Galvao, and Montes-Rojas (2012). (See also Koenker, 2004 and Harding and Lamarche, 2019). In contrast, the penalized quantile regression estimator (PQR) reduces the bias and RMSE for small values of λ_T. The evidence shows that small increases of the tuning parameter lead to substantial improvements in both the bias profile and the RMSE.

S.3.2. **Inference.** We now turn our attention to the performance of tests using the bootstrap. To this end, Table S.1 reports empirical rejection frequencies for the null hypothesis $H_0 : \beta_0 = 1 + \zeta F_u(0.5)^{-1}$. As in Table 4.1, we consider different sample sizes $N \in \{100, 200\}$ and $T \in \{5, 10\}$, different distributions F_u, and different assumptions on α_i. We report results using two different approaches. The cross-sectional pairs bootstrap (CS) samples over i with replacement, keeping the entire block of time series observations. The wild bootstrap is implemented as discussed in Section 2.3. We first obtain residuals $\hat{u}_{it}$ using the penalized quantile regression estimator. The estimator (2.6) is labeled 'WB1' and the estimator (2.7) is labeled 'WB2'. As in the case of the wild bootstrap

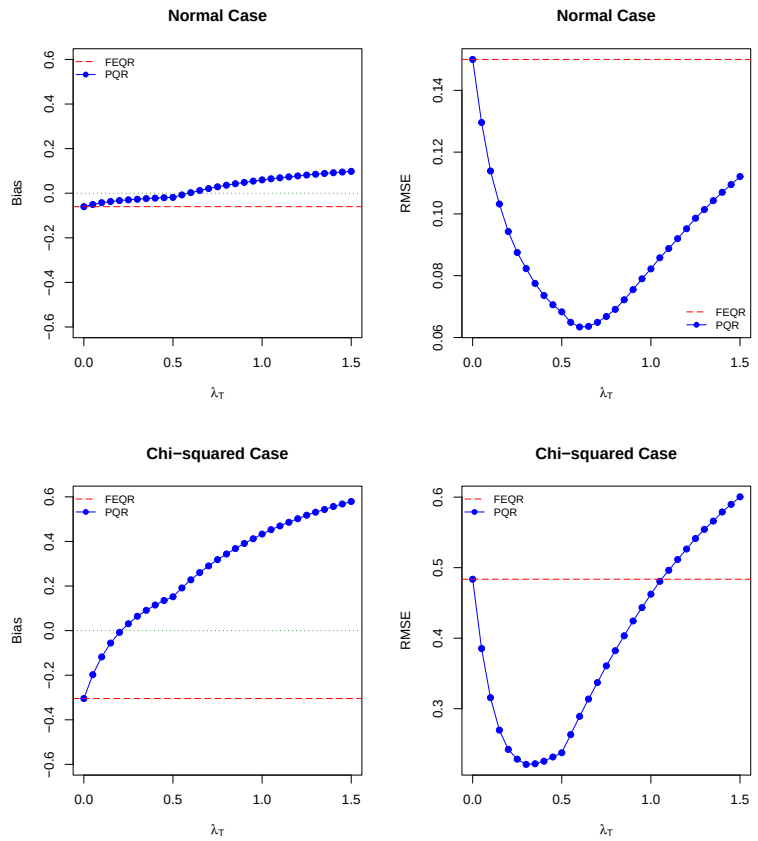

FIGURE S.1. *Small sample performance of the fixed effects (FE) and penalized quantile regression (PQR) in a location-scale shift model.*

estimator proposed by Feng, He, and Hu (2011), a finite sample correction is recommended. 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 first columns report results based on bootstrap critical values obtained from the distribution of $\sqrt{NT}(\beta^* - \hat{\beta})$, where $\hat{\lambda}_T$ is obtained as in Table 4.1. The last columns report results obtained using bootstrap standard errors, which are denoted by $\text{se}(\beta^*)$. In this case, the statistic is $|\hat{\beta} - \beta_0|/\text{se}(\beta^*)$ and it is compared to $\Phi^{-1}(1-\alpha/2)$. The theoretical size of the tests is equal to 5%. As it can be seen in the upper block of Table S.1, the wild bootstrap procedure tends to produce empirical sizes that are closer to the nominal values. The lower panels of Table S.1 show results for a DGP when the error term is distributed as t_3 and χ^2_3 and offer similar conclusions. We do not observe significant differences between probabilities estimated by bootstrap critical values or bootstrap standard errors.

		Bootstrap Critical Values						Bootstrap Standard Errors					
		Method:			Method:			Method:			Method:		
N	T	CS	WB1	WB2	CS	WB1	WB2	CS	WB1	WB2	CS	WB1	WB2
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	0.008	0.052	0.048	0.038	0.050	0.041	0.018	0.038	0.041	0.088	0.043	0.042
100	10	0.003	0.040	0.041	0.033	0.053	0.044	0.009	0.035	0.040	0.086	0.049	0.047
200	5	0.004	0.041	0.039	0.023	0.030	0.033	0.012	0.038	0.038	0.069	0.029	0.033
200	10	0.004	0.036	0.039	0.023	0.041	0.040	0.008	0.036	0.038	0.067	0.036	0.042
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	0.042	0.062	0.061	0.045	0.064	0.064	0.087	0.049	0.050	0.094	0.061	0.061
100	10	0.041	0.049	0.047	0.046	0.053	0.053	0.091	0.040	0.041	0.105	0.053	0.051
200	5	0.038	0.052	0.048	0.028	0.037	0.037	0.079	0.038	0.042	0.073	0.031	0.033
200	10	0.032	0.041	0.040	0.034	0.037	0.039	0.085	0.038	0.040	0.100	0.037	0.038
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	0.004	0.043	0.039	0.045	0.052	0.042	0.022	0.036	0.035	0.104	0.051	0.043
100	10	0.004	0.035	0.034	0.031	0.039	0.040	0.009	0.034	0.031	0.085	0.040	0.042
200	5	0.010	0.030	0.033	0.039	0.042	0.038	0.019	0.025	0.030	0.102	0.042	0.044
200	10	0.004	0.031	0.030	0.035	0.042	0.040	0.009	0.034	0.032	0.085	0.038	0.041
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	0.032	0.038	0.035	0.062	0.055	0.054	0.065	0.038	0.040	0.100	0.054	0.054
100	10	0.046	0.041	0.044	0.054	0.057	0.056	0.090	0.042	0.043	0.105	0.051	0.054
200	5	0.032	0.026	0.029	0.052	0.052	0.047	0.065	0.033	0.030	0.097	0.052	0.050
200	10	0.044	0.036	0.037	0.043	0.033	0.033	0.090	0.038	0.037	0.101	0.033	0.033
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	0.022	0.047	0.045	0.040	0.062	0.062	0.051	0.046	0.047	0.107	0.070	0.070
100	10	0.024	0.061	0.066	0.042	0.068	0.068	0.056	0.059	0.064	0.100	0.063	0.064
200	5	0.034	0.049	0.052	0.046	0.046	0.043	0.059	0.044	0.045	0.113	0.062	0.061
200	10	0.030	0.064	0.068	0.037	0.043	0.043	0.059	0.065	0.067	0.079	0.044	0.044
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	0.046	0.055	0.055	0.052	0.066	0.064	0.076	0.057	0.058	0.095	0.074	0.070
100	10	0.037	0.056	0.056	0.049	0.068	0.069	0.086	0.053	0.054	0.114	0.067	0.069
200	5	0.061	0.056	0.057	0.053	0.057	0.053	0.098	0.068	0.067	0.107	0.074	0.074
200	10	0.043	0.061	0.062	0.038	0.046	0.043	0.113	0.060	0.061	0.097	0.045	0.045

TABLE S.1. *Empirical rejection probabilities of* $H_0 : \beta_0(0.5) = 1 + \zeta F_u(0.5)^{-1}$. *CS denotes cross-sectional pairs bootstrap, WB1 denotes wild bootstrap estimator* (2.6), *and WB2 wild bootstrap estimator* (2.7).

REFERENCES

CHAO, S.-K., S. VOLGUSHEV, AND G. CHENG (2017): "Quantile Processes for Semi and Nonparametric Regression," *Electronic Journal of Statistics*, 11, 3272–3331.

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.

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

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

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

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. Wiley, New York.