Asymptotics and bootstrap inference for panel quantile regression models with fixed effects

Kengo Kato* Antonio F. Galvao, Jr.† Gabriel V. Montes-Rojas‡

February 1, 2010

Abstract

This paper studies panel quantile regression models with fixed effects. We formally establish sufficient conditions for consistency and asymptotic normality of the quantile regression estimator when the number of individuals, n, and the number of time periods, T, jointly go to infinity. The estimator is shown to be consistent under similar conditions to those found in the nonlinear panel data literature. Nevertheless, due to the non-smoothness of the criterion function, we had to impose a more restrictive condition on T to prove asymptotic normality than that usually found in the literature. We also examine the practical ability of bootstrap procedures for inference in quantile regression models for panel data. The finite sample performance of the estimator and the bootstrap procedures are evaluated by Monte Carlo simulations.

Key words: asymptotics, bootstrap, fixed effects, panel data, quantile regression.

JEL Classification: C13, C21, C23.

1 Introduction

Quantile regression (QR) for panel data has attracted considerable interest in both the theoretical and applied literatures. It allows us to explore a range of conditional quantiles, thereby exposing a variety of forms of conditional heterogeneity, and to control for unobserved individual effects. Controlling for individual heterogeneity via fixed effects, while

*Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. E:mail: kkato@hiroshima-u.ac.jp

†Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201. E-mail: agalvao@uwm.edu

‡Department of Economics, City University London, D306 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK. Email: Gabriel.Montes-Rojas.1@city.ac.uk

exploring heterogeneous covariate effects within the QR framework, offers a more flexible approach to the analysis of panel data than that afforded by the classical Gaussian fixed and random effects estimation.

This paper focuses on the fixed effect estimation of the common parameters in a QR model with individual effects. We refer to the resulting estimator as the fixed-effect quantile regression (FE-QR) estimator. Unfortunately, the FE-QR estimator is subject to the incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000, for a review) and will be inconsistent if the number of individuals n goes to infinity while the number of time periods T is fixed. It is important to note that, in contrast to mean regression, there is no general transformation that can suitably eliminate the individual effects in the QR model. Therefore, no clever way is known to sidestep the incidental parameters problem in the QR case.

The incidental parameters problem has been extensively studied in the recent nonlinear panel data literature. Among them, Hahn and Newey (2004) studied the maximum likelihood estimation of a general nonlinear panel data model with fixed effects. They showed that the maximum likelihood estimator (MLE) has a limiting normal distribution with a bias in the mean when n and T grow at the same rate, and proposed several bias correction methods to the MLE. Note that since they assumed that likelihood functions are smooth, while the criterion function of QR is not, their results are not directly applicable to the QR case.

Koenker (2004) introduced a novel approach to the estimation of a QR model for panel data. He argued that shrinking the individual parameters towards a common value improves the performance of the common parameters' estimates, and proposed a penalized estimation method where the individual parameters are subject to the ℓ_1 penalty. He also studied the asymptotic properties of the (unpenalized) FE-QR estimator and claimed that it is asymptotically normal provided that $n^a/T \to 0$ for some $a > 0$. We provide an alternative formal and rigorous approach that offers a clearer understanding of the asymptotic properties of the FE-QR estimator than that in his Theorem 1.

The goals of this paper are twofold. First, we study the asymptotic properties of the FE-QR estimator when n and T jointly go to infinity and formally establish sufficient conditions for consistency and asymptotic normality of the estimator. We show that the FE-QR estimator is consistent under similar conditions to those found in the nonlinear panel data literature. However, we need a more restrictive condition on T (i.e., $n^2(\log n)^3/T \to 0$) to prove asymptotic normality of the estimator than that found in the literature. This reflects the fact that the order of the remainder term in the Bahadur representation of the FE-QR estimator is $O_p\{T^{-3/4}(\log n)^{3/4}\}$, while that of the smooth MLE is $O_p(T^{-1})$. The slower convergence rate of the remainder term is due to the non-smoothness of the scores.

From a methodological point of view, the proof of asymptotic normality of the FE-QR estimator is of independent interest. Because of the non-differentiability of the criterion function, the stochastic expansion technique of Li, Lindsay, and Waterman (2003) is no longer applicable to the asymptotic analysis of the FE-QR estimator. Instead, we adapt the Pakes and Pollard (1989) approach for proving asymptotic normality of the estimator. In addition, we make use of Talagrand's (1994,1996) inequalities available from the empirical process literature to establish the convergence rate of the remainder term in the Bahadur representation of the FE-QR estimator. These inequalities significantly simplify the proof.

The second goal of this paper is to examine the practical ability of bootstrap procedures to construct confidence intervals for the QR model with fixed effects. Due to the specific nature of panel data, there are three possibilities for bootstrap resampling, namely, cross-sectional resampling, temporal resampling and cross-sectional & temporal resampling. We use Monte Carlo simulations to study the finite sample performance of the FE-QR estimator and the bootstrap procedures. The simulation study highlights some cases where the FE-QR estimator has large bias in panels with small T. The results show that the cross-sectional resampling outperforms the others and the empirical levels from it approximate well the nominal level when T is of moderate size, but none of the procedures may provide accurate confidence intervals in some cases when T is small relative to n.

We now review the literature related to this paper. Lamarche (2006) studied Koenker's penalization method and discussed an optimal choice of the tuning parameter. Canay (2008) adopted a Doksum (1974) type formulation of a QR model for panel data, and proposed a two-step estimator of the common parameters. The difference is that in his model, each individual effect is not allowed to change across quantiles. Graham, Hahn, and Powell (2009) showed that when $T = 2$ and the explanatory variables are independent of the error term, the FE-QR estimator does not suffer from the incidental parameters problem. However, their argument does not apply to the general case. Rosen (2009) addressed a set identification problem of the common parameters when T is fixed. Chernozhukov, Fernandez-Val, and Newey (2009) considered identification and estimation of the quantile structural function defined in Imbens and Newey (2009) of a nonseparable panel model with discrete explanatory variables. They studied bounds of the quantile structural function when T is fixed and the asymptotic behavior of the bounds when T goes to infinity. Note that the quantile structural function is not equivalent to the conditional quantile function which we study in this paper.

This paper is organized as follows. In Section 2, we introduce a QR model with fixed effects and discuss the asymptotic properties of the FE-QR estimator. In Section 3, we introduce three types of bootstrap procedures for approximating the finite sample distribution of the estimator. In Section 4, we report a simulation study for assessing the finite sample performance of the FE-QR estimator and the bootstrap procedures described in Section 3. In Section 5, we present some discussion on the paper. We relegate the proofs of the theorems to the Appendix.

2 Quantile regression with fixed effects

2.1 The model

In this paper, we consider a QR model with individual effects

$$Q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_i^*(\tau)) = \alpha_i^*(\tau) + \boldsymbol{x}_{it}'\boldsymbol{\beta}_0(\tau), \tag{2.1}$$

where y_{it} is a response variable, $\boldsymbol{x}_{it}$ is a p dimensional vector of explanatory variables and $Q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_i^*(\tau))$ is the conditional τ-quantile of y_{it} given $\boldsymbol{x}_{it}$ and $\alpha_i^*(\tau)$. In model (2.1), each $\alpha_i^*(\tau)$ is intended to capture some individual specific source of variability or unobserved heterogeneity that is not adequately controlled for by the explanatory variables. We make no parametric assumption on the relationship between $\alpha_i^*(\tau)$ and $\boldsymbol{x}_{it}$. Throughout the paper, the number of individuals is denoted by n and the number of time periods is denoted by $T = T_n$ which depends on n. Usually, we omit the subscript n of T_n.

As in Koenker (2004), we treat each individual effect as a parameter to be estimated and consider the estimator

$$(\hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\tau)) := \arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_\tau(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta}), \tag{2.2}$$

where $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)'$ and $\rho_\tau(u) := \{\tau - I(u \leq 0)\}u$ is the check function of Koenker and Bassett (1978). Note that $\boldsymbol{\alpha}$ implicitly depends on n. We call $\hat{\boldsymbol{\beta}}(\tau)$ the fixed effects quantile regression (FE-QR) estimator of $\boldsymbol{\beta}_0(\tau)$. The optimization for solving (2.2) can be very large depending on n and T. However, as Koenker (2004) observed, in typical applications, the design matrix is very sparse. Standard sparse matrix storage schemes only require the space for the non-zero elements and their indexing locations. This considerably reduces the computational effort and memory requirements.

It is important to note that in the QR model, there is no general transformation that can suitably eliminate the individual effects. This intrinsic difficulty has been recognized by Abrevaya and Dahl (2008), among others, and was clarified by Koenker and Hallock (2000). They remarked that "Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects. (p.19)"

2.2 Asymptotic theory

In this section, we investigate the asymptotic properties of the FE-QR estimator. In order to avoid notational complication, we suppress the dependence on τ throughout the section. For example, we simply write α_i^* for $\alpha_i^*(\tau)$. In the asymptotic analysis, as in Hahn and Newey (2004) and Fernandez-Val (2005), we pick a single realization $(\alpha_{10}, \alpha_{20}, \dots)$ of $(\alpha_1^*, \alpha_2^*, \dots)$ and treat each α_{i0} as a fixed true parameter. From now on, we suppose that every argument is conditional on $\alpha_i^* = \alpha_{i0}$ for each individual i. Put $\boldsymbol{\alpha}_0 := (\alpha_{10}, \dots, \alpha_{n0})'$.

We first consider the consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. We say that $\hat{\boldsymbol{\alpha}}$ is weakly consistent if $\hat{\alpha}_i$ converges in probability to α_{i0} uniformly over $1 \leq i \leq n$, that is, $\max_{1\leq i \leq n} |\hat{\alpha}_i - \alpha_{i0}| \xrightarrow{p} 0$. Now, we introduce some regularity conditions that ensure the consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$.

(A1) $\{(y_{it}, \boldsymbol{x}_{it}), t \geq 1\}$ is independent and identically distributed (i.i.d.) for each fixed i and independent across i.

(A2) $\sup_{i\geq 1} \mathrm{E}[\|\boldsymbol{x}_{i1}\|^{2s}] < \infty$ for some real $s \geq 1$.

The distribution of $(y_{it}, \boldsymbol{x}_{it})$ is allowed to depend on i. Put $u_{it} := y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_0$. Condition (A1) implies that $\{(u_{it}, \boldsymbol{x}_{it}), t \geq 1\}$ is i.i.d. for each fixed i and independent across i. Let $F_i(u|\boldsymbol{x})$ denote the conditional distribution function of u_{it} given $\boldsymbol{x}_{it} = \boldsymbol{x}$. We assume that $F_i(u|\boldsymbol{x})$ has density $f_i(u|\boldsymbol{x})$. Let $f_i(u)$ denote the marginal density of u_{it}.

(A3) For each $\delta > 0$,

$$\epsilon_\delta := \inf_{i\geq 1} \inf_{|\alpha|+\|\boldsymbol{\beta}\|_1=\delta} \mathrm{E}\left[\int_0^{\alpha+\boldsymbol{x}_{i1}'\boldsymbol{\beta}} \{F_i(s|\boldsymbol{x}_{i1}) - \tau\} ds\right] > 0, \tag{2.3}$$

where $\|\cdot\|_1$ stands for the ℓ_1 norm.

Condition (A1) is the same as Condition 1 (i) in Fernandez-Val (2005). We exclude temporal dependence to simplify the technical proofs. Condition (A2) corresponds to the moment condition in Fernandez-Val (2005, p.12). Condition (A3) is a high level condition for identification of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and corresponds to Condition 3 in Hahn and Newey (2004). In fact, it is sufficient for consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ that (2.3) is satisfied for any sufficiently small $\delta > 0$. Recall that $F_i(0|\mathbf{x}_{i1}) = \tau$. Under suitable integrability conditions, the expectation in (2.3) can be expanded as $(\alpha, \boldsymbol{\beta}')\Omega_i(\alpha, \boldsymbol{\beta}')' + o(\delta^2)$ for $|\alpha| + \|\boldsymbol{\beta}\|_1 = \delta$ uniformly over $i \geq 1$ as $\delta \to 0$, where $\Omega_i := \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})(1, \boldsymbol{x}_{i1}')(1, \boldsymbol{x}_{i1}')']$. If the minimum eigenvalue of Ω_i is bounded away from zero uniformly over $i \geq 1$, there exists a positive constant δ_0 such that for $0 < \delta \leq \delta_0$, (2.3) is satisfied.

Theorem 2.1. *Assume that $n/T^s \to 0$ as $n \to \infty$, where s is given in condition (A2). Then, under conditions (A1)-(A3), $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is weakly consistent.*

Remark 2.1.

(a) Theorem 2.1 is not covered by Hahn and Newey (2004) and Fernandez-Val (2005) because they assumed that the parameter spaces of α_{i0} and $\boldsymbol{\beta}_0$ are compact. In our problem, due to the convexity of the criterion function, we can remove the compactness assumption of the parameter spaces.

(b) The condition on T in Theorem 2.1 is the same as that in Theorems 1-2 of Fernandez-Val (2005). If $\sup_{i\geq 1}\|\boldsymbol{x}_{i1}\| \leq M$ (a.s.) for some positive constant M, then the conclusion of the theorem holds when $\log n/T \to 0$ as $n \to \infty$. See Remark A.1 in the Appendix for details.

Next, we derive the limiting distribution of $\hat{\boldsymbol{\beta}}$. To do this, we present another set of conditions.

(B1) There exists a constant M such that $\sup_{i\geq 1}\|\boldsymbol{x}_{i1}\| \leq M$ (a.s.).

(B2) (a) For each i, $f_i(u|\boldsymbol{x})$ is continuously differentiable with respect to u for each $\boldsymbol{x}$ and let $f_i^{(1)}(u|\boldsymbol{x}) := \partial f_i(u|\boldsymbol{x})/\partial u$; (b) there exist constants A_0 and A_1 such that $f_i(u|\boldsymbol{x}) \leq A_0$ and $|f_i^{(1)}(u|\boldsymbol{x})| \leq A_1$ uniformly over $(u,\boldsymbol{x})$ and $i \geq 1$; (c) $f_i(0)$ is bounded from below by some positive constant independent of i.

(B3) Put $\boldsymbol{\gamma}_i := \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}]/f_i(0)$ and $\Gamma_n := n^{-1}\sum_{i=1}^n \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}(\boldsymbol{x}_{i1}' - \boldsymbol{\gamma}_i')]$. (a) Γ_n is nonsingular for each n, and the limit $\Gamma := \lim_{n\to\infty}\Gamma_n$ exists and is nonsingular; (b) the limit $V := \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathrm{E}[(\boldsymbol{x}_{i1}-\boldsymbol{\gamma}_i)(\boldsymbol{x}_{i1}-\boldsymbol{\gamma}_i)']$ exists and is nonsingular.

Condition (B1) is assumed in Koenker (2004). This condition is used to ensure the "asymptotic" first order condition displayed in equation (A.7) in the proof of Theorem 2.2. Condition (B2) imposes some restrictions on the conditional density and seems standard in the QR literature. Compare condition (ii) of Theorem 3 in Angrist, Chernozhukov, and Fernandez-Val (2006). Condition (B3) is concerned with the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. Condition (B3) (a) implies that the minimum eigenvalue of Γ_n is bounded away from zero uniformly over $n \geq 1$.

Theorem 2.2. *Assume that $T = O(n^r)$ for some $r > 0$ and $\log n/T \to 0$ as $n \to \infty$. Then, under conditions (A1), (A3) and (B1)-(B3), $\hat{\boldsymbol{\beta}}$ admits the representation*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) = \Gamma_n^{-1}\left\{\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\{\tau - I(u_{it} \leq 0)\}(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)\right\} + R_n, \tag{2.4}$$

where R_n is $O_p\{T^{-3/4}(\log n)^{3/4}\}$. If in addition $n^2(\log n)^3/T \to 0$ as $n \to \infty$, we have

$$\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\{\mathbf{0}, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}\}. \tag{2.5}$$

Remark 2.2.

(a) [Relation to Hahn and Newey (2004)] Equations (10) and (17) in Hahn and Newey (2004) show that the MLE of the common parameters for smooth likelihood functions admits the representation

$$\hat{\theta} - \theta_0 = \left(\frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i \right)^{-1} \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} U_{it} \right) + \frac{1}{2T} \theta^{\epsilon\epsilon}(0) + \frac{1}{6T^{3/2}} \theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}), \tag{2.6}$$

where $\hat{\theta}, \theta_0, \mathcal{I}_i, U_{it}, \theta^{\epsilon\epsilon}(\cdot)$ and $\theta^{\epsilon\epsilon\epsilon}(\cdot)$ are defined in Hahn and Newey (2004) and $\tilde{\epsilon}$ is in $[0, T^{-1/2}]$. Under suitable regularity conditions, $\theta^{\epsilon\epsilon}(0)$ is $O_p(1)$ and $\theta^{\epsilon\epsilon\epsilon}(\epsilon)$ is $O_p(1)$ uniformly over $\epsilon \in [0, T^{-1/2}]$, which implies that the last two terms in the right hand side of equation (2.6) are $O_p(T^{-1})$ and $O_p(T^{-3/2})$, respectively.[1] This difference is due to the non-smoothness of the scores. Hahn and Newey (2004) assumed that the scores are sufficiently smooth with respect to the parameters. On the other hand, the scores for problem (2.2), which are formally defined in Appendix A.2, are not differentiable. See Remark A.2 for the technical reason why the (non-)smoothness of the scores is essential for the convergence rate of the remainder term.

(b) [Relation to He and Shao (2000)] He and Shao (2000) studied a general M-estimation with diverging number of parameters which allows for non-smooth criterion functions. It is interesting to note that their Corollary 3.2 shows that the smoothness of scores is crucial for the growth condition of the number of parameters in asymptotic distribution theory of M-estimators. However, it should be pointed out that our Theorem 2.2 is not derived from their result because of the specific nature of the panel data problem. The formal problem to apply their result is that the convergence rate of $\hat{\alpha}_i$ is different from that of $\hat{\boldsymbol{\beta}}$. To avoid this, make a reparametrization $\boldsymbol{\theta} = (n^{-1/2}\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ and put $\boldsymbol{z}_{it} := (n^{1/2}\boldsymbol{e}_i', \boldsymbol{x}_{it}')'$, where $\boldsymbol{e}_i$ is the i-th unit vector in $\mathbb{R}^n$. Then, the current problem is under the framework of He and Shao (2000) with $x_i = (y_{it}, \boldsymbol{z}_{it}), m = (n+p), p = (n+p), n = nT, \theta = \boldsymbol{\theta}$ and $\psi(x_i, \theta) = \{\tau - I(y_{it} \leq \boldsymbol{z}_{it}'\boldsymbol{\theta})\}\boldsymbol{z}_{it}$.[2] Although conditions (C0)-(C3) may be achieved in this case, it is difficult to obtain a tight bound of $A(n,m)$ in conditions (C4) and (C5) of their paper. If we use the same reasoning as in Lemma 2.1 of He and Shao (2000), $A(n,m)$ is bounded by a constant times $n^{3/2}T^{1/2}$ (in our notation), but if we use this bound, the condition on T implied by Theorem 2.2 of He and Shao (2000) will be such that $n^3(\log n)^2/T \to 0$.

[1] In fact, Hahn and Newey (2004) showed that $\theta^{\epsilon\epsilon}(0)$ converges in probability to some constant vector, which will contribute to the bias in the asymptotic distribution when n and T grow at the same rate.

[2] The left hand sides correspond to the notation of He and Shao (2000) and the right hand sides correspond to our notation.

(c) It should be pointed out that although the above rate of the remainder term is the best one (up to $\log n$) that we could achieve, there might be a room for improvement on the rate. Nevertheless, a substantial improvement on the rate of the remainder term (if possible) is expected to be quite complicated, and we leave it for future research.

(d) The proof of Theorem 2.2 is of independent interest. In contrast to estimators with smooth criterion functions which have been studied in the literature such as Li, Lindsay, and Waterman (2003), Hahn and Newey (2004) and Fernandez-Val (2005), the Taylor-series methods of asymptotic distribution theory do not apply to the FE-QR estimator, which greatly complicates the analysis of its asymptotic distributional properties. The difficulty is partly explained by the fact that, as Hahn and Newey (2004) observed, the first order asymptotic behavior of the (smooth) MLE of the common parameters can be affected by the second order behavior of the estimators of the individual parameters, while the second order behavior of QR estimators is non-standard and rather complicated (Arcones, 1998; Knight, 1998).

(e) The proof proceeds as follows. It is based on the method of Pakes and Pollard (1989), but requires some extra efforts. The first step is to obtain certain representations of $\hat{\alpha}_i - \alpha_{i0}$ by expanding the first n elements of the scores. Plugging them into the expansion of the last p elements of the scores, we obtain a representation of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ (see (A.5)). The remaining task is to evaluate the remainder terms in the representation of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$, which corresponds to establishing the stochastic equicontinuity condition in Pakes and Pollard (1989). However, since the number of parameters goes to infinity as $n \to \infty$, the "standard" empirical process argument such as that displayed in their paper will not suffice to show this. In order to establish the convergence rate of the remainder terms, we make use of moment and exponential inequalities for general empirical processes developed in Talagrand (1994, 1996), which significantly simplify the proof. We summarize these inequalities in Appendix B.

(f) [Relation to Koenker (2004)] Koenker (2004) claimed asymptotic normality of the FE-QR estimator under similar conditions to ours except that he assumed that $n^a/T \to 0$ for some $a > 0$. We believe that our proof of asymptotic normality offers a clearer understanding of the asymptotic properties of the FE-QR estimator than that in his Theorem 1. Actually, in his proof, a formal proof for $\sqrt{nT}$-consistency of $\hat{\boldsymbol{\beta}}$ is not offered, and a justification for the second expression of R_{mn} in p.82 when n and m (in his notation) jointly go to infinity is not presented.

3 Bootstrap inference

The main concern of this section is the application of bootstrap procedures to the problem of constructing confidence intervals for the common parameters of the QR model with fixed effects. Theorem 2.2 suggests that in order to use the normal approximation T should be larger than n^2, which seems restrictive in practice. Even in cases where we can use the normal approximation, the asymptotic covariance matrix is not easy to estimate since it depends on the conditional density. We consider the bootstrap as a heuristic method to approximate the finite sample distribution of the FE-QR estimator and to investigate the finite sample performance of the bootstrap through simulations. The simulation study appears in the next section.

Traditionally, the bootstrap has been successfully employed to construct confidence intervals for parameters of QR models for cross-section data.[3] However, it is only recently that the properties of the bootstrap for panel data has been studied. Cameron and Trivedi (2005) discussed resampling methods for panel data when n is large but T is assumed small. Kapetanios (2008) studied the bootstrap for a linear panel data model where resampling occurs in both the cross-section and time dimensions. Goncalves (2008), allowing for both temporal and cross-sectional dependence, studied the moving block bootstrap for a linear panel data model where resampling occurs only in the time dimension.

In this section, following Kapetanios (2008), we consider three different methods of bootstrap resampling, namely, *cross-sectional resampling*, *temporal resampling* and *cross-sectional & temporal resampling*. Put $\boldsymbol{y}_{i\cdot} = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{X}_{i\cdot} = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})'$, $\boldsymbol{y}_{\cdot t} = (y_{1t}, \ldots, y_{nt})'$ and $\boldsymbol{X}_{\cdot t} = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{nt})'$. The cross-sectional resampling consists of resampling $(\boldsymbol{y}_{i\cdot}, \boldsymbol{X}_{i\cdot})$ with replacement from the cross-section dimension with probability $1/n$. The temporal resampling consists of resampling $(\boldsymbol{y}_{\cdot t}, \boldsymbol{X}_{\cdot t})$ with replacement from the temporal dimension for each individual with probability $1/T$. The cross-sectional & temporal resampling involves both the cross-sectional and temporal resamplings. It first resamples $(\boldsymbol{y}_{i\cdot}, \boldsymbol{X}_{i\cdot})$ from the cross-sectional dimension with probability $1/n$ and then resamples the constructed cross-sectional units from the time dimension with probability $1/T$.

Having obtained the resampled data $\{(y_{it}^*, \boldsymbol{x}_{it}^*), i = 1, \ldots, n; t = 1, \ldots, T\}$, the bootstrap estimate is computed as

$$(\hat{\boldsymbol{\alpha}}^*(\tau), \hat{\boldsymbol{\beta}}^*(\tau)) := \arg\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} w_{it} \rho_\tau (y_{it} - \alpha_i - \boldsymbol{x}_{it}' \boldsymbol{\beta}),$$

where w_{it} is the number of times that $(y_{it}, \boldsymbol{x}_{it})$ is "redrawn" from the original sample. The distribution of $\hat{\boldsymbol{\beta}}^*(\tau) - \hat{\boldsymbol{\beta}}(\tau)$ conditional on the sample would approximate the finite sample distribution of $\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)$. In practice, we repeat these steps a large number of times.

[3]For example, see Buchinsky (1995), Hahn (1995) and Horowitz (1998).

Then, the confidence intervals for $\boldsymbol{\beta}_0(\tau)$ can be obtained by computing the quantiles of the empirical distribution of $\hat{\boldsymbol{\beta}}^*(\tau) - \hat{\boldsymbol{\beta}}(\tau)$ conditional on the sample.

There is an alternative way to compute $\hat{\boldsymbol{\beta}}^*(\tau)$. $\hat{\boldsymbol{\beta}}^*(\tau)$ can be also defined as

$$(\hat{\boldsymbol{c}}^*, \hat{\boldsymbol{\beta}}^*(\tau)) := \arg\min_{\boldsymbol{c},\boldsymbol{\beta}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_\tau(y_{it}^* - c_i - \boldsymbol{x}_{it}^{*\prime}\boldsymbol{\beta}).$$

Here, $\hat{\boldsymbol{c}}^*$ is not necessarily equal to $\hat{\boldsymbol{\alpha}}^*(\tau)$ when the cross-sectional resampling is included but it is not difficult to see that the two definitions of $\hat{\boldsymbol{\beta}}^*(\tau)$ are equivalent.

4 Monte Carlo

In this section, we report a simulation study to investigate the finite sample performance of the bootstrap procedures described in Section 3. Two simple versions of model (2.1) are considered in the simulation study:

1. Location shift model: $y_{it} = \eta_i + x_{it} + \epsilon_{it}$;
2. Location-scale shift model: $y_{it} = \eta_i + x_{it} + (1 + 0.2x_{it})\epsilon_{it}$,

where $x_{it} = 0.3\eta_i + z_{it},\ z_{it} \sim$ i.i.d. χ_3^2, $\eta_i \sim$ i.i.d. $U[0,1]$ and $\epsilon_{it} \sim$ i.i.d. F with $F = N(0,1)$ or χ_3^2. In the location shift model, $\alpha_i^*(\tau) = \eta_i + F^{-1}(\tau)$ and $\beta_0(\tau) = 1$, while in the location-scale shift model, $\alpha_i^*(\tau) = \eta_i + F^{-1}(\tau)$ and $\beta_0(\tau) = 1 + 0.2F^{-1}(\tau)$. Finally, we consider cases where $n \in \{25, 50\}$, $T \in \{5, 10, 50\}$ and $\tau \in \{0.25, 0.50, 0.75\}$.

4.1 Bias

The performance of the FE-QR estimator is evaluated by its bias and standard deviation. The number of Monte Carlo repetitions is $10,000$. Tables 1 and 2 report the results for the location shift and location-scale shift models, respectively. For the median, the results are in line with those of Koenker (2004), where in both models the FE-QR estimator has small bias and standard deviation in small samples. However, there are noticeable differences for the first and third quartiles. In the location shift model, the bias is very small in every case and the standard deviation decreases monotonically as either n or T increases. In the location-scale shift model, however, both bias and standard deviation are large for small T. In particular, the bias is moderate in the χ_3^2 case for the third quartile and $T = 5, 10$. As expected, the bias disappears as T increases. These results suggest that the FE-QR estimator performs well in small samples for the location shift model but may have a large bias for the location-scale shift model where the quantile of interest is evaluated at an associated low density (i.e., $F = \chi_3^2$ and $\tau = 0.75$ case) when T is small.

4.2 Bootstrap

The performance of confidence intervals constructed from bootstrap estimates is evaluated by the (empirical) rejection rate and length. The rejection rate is defined as the proportion of cases that the true parameter is outside the confidence intervals. If the bootstrap procedure correctly approximates the finite sample distribution of $\hat{\beta}(\tau) - \beta_0(\tau)$, the rejection rate should be close to the nominal size. The number of bootstrap repetitions is 500 for each Monte Carlo repetition. Because of the computational burden, the number of Monte Carlo repetitions is 500. The nominal size is set to 0.1.

Table 3 reports the results for the location shift model. The cross-sectional resampling performs the best. In this case, rejection rates are close to 0.1 except for a few cases. The temporal resampling gives rejection rates well below the nominal size in all cases. Finally, confidence intervals based on the cross-sectional & temporal resampling are excessively wide and as a result give very small rejection rates.

Table 4 reports the results for the location-scale shift model. The cross-sectional resampling again performs the best but its performance is worse than in the linear location shift model. In this case, there are considerable differences among quantiles. In the normal case, rejection rates for the median are close to the nominal size in all cases. However, for the first and third quartiles, rejection rates are double the nominal size when n/T is large (the lower diagonal part in each box). In the χ^2 case, rejection rates are acceptable when $\tau \in \{0.25, 0.5\}$ but double the nominal size when $\tau = 0.75$ and n/T is large. This is possibly due to the large bias found in Table 2. The other resampling procedures have a poorer performance than the cross-sectional resampling. The temporal resampling has rejection rates close to the nominal size only when $\tau = 0.5$ and $T = 50$. Not surprisingly, the temporal resampling is unreliable when T is small. The performance of the cross-sectional & temporal resampling is sensitive to the setting.

In summary, the Monte Carlo results suggest that (a) the cross-sectional resampling is recommended among three resampling procedures and performs well when both n and T are of moderate sizes; however, (b) none of the resampling procedures may not be accurate in some cases (such as the location-scale shift model with $\tau \in \{0.25, 0.75\}$) when n/T is large.

5 Discussion

In this paper, we have studied the asymptotic properties of the FE-QR estimator and examined the practical ability of three bootstrap procedures by means of simulations. There remain several issues to be investigated. It is an open question whether the convergence rate of the remainder term in (2.4) can be improved (it is clear that the rate can not be faster than $O_p(T^{-1})$). Investigation into this direction will require a sophisticated probabilistic argument. Also, the simulation result suggests that some modification is desirable to improve the performance of the FE-QR estimator under small T relative to n. The authors plan to pursue this issue in a separate paper.

A Proofs

A.1 Proof of Theorem 2.1

Put $\mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) := T^{-1}\sum_{t=1}^{T} \rho_\tau(y_{it} - \alpha_i - \boldsymbol{x}_{it}'\boldsymbol{\beta})$ and $\Delta_{ni}(\alpha_i, \boldsymbol{\beta}) := \mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)$. For each $\delta > 0$, define $B_i(\delta) := \{(\alpha, \boldsymbol{\beta}) : |\alpha - \alpha_{i0}| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 \le \delta\}$ and $\partial B_i(\delta) := \{(\alpha, \boldsymbol{\beta}) : |\alpha - \alpha_{i0}| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1 = \delta\}$.

Proof of Theorem 2.1. We divide the proof into two steps.

Step 1. We first prove $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. Fix any $\delta > 0$. For each $(\alpha_i, \boldsymbol{\beta}) \notin B_i(\delta)$, define $\tilde{\alpha}_i = r_i\alpha_i + (1 - r_i)\alpha_{i0}$, $\tilde{\boldsymbol{\beta}} = r_i\boldsymbol{\beta} + (1 - r_i)\boldsymbol{\beta}_0$, where $r_i = \delta/(|\alpha_i - \alpha_{i0}| + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_1)$. Note that $r_i \in (0, 1)$ and $(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}) \in \partial B_i(\delta)$. Because of the convexity of the criterion function, we have

$$\begin{aligned} r_i\{\mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)\} &\ge \mathbb{M}_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0) \\ &= \{\mathrm{E}[\Delta_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}})]\} + \{\Delta_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}}) - \mathrm{E}[\Delta_{ni}(\tilde{\alpha}_i, \tilde{\boldsymbol{\beta}})]\}. \quad \text{(A.1)} \end{aligned}$$

Use the identity of Knight (1998) to obtain

$$\mathrm{E}[\Delta_{ni}(\alpha_i, \boldsymbol{\beta})] = \mathrm{E}\left[\int_0^{(\alpha_i - \alpha_{i0}) + \boldsymbol{x}_{i1}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)} \{F_i(s|\boldsymbol{x}_{i1}) - \tau\} ds\right].$$

From condition (A3), the first term in the right hand side of equation (A.1) is greater than or equal to ϵ_δ for all $1 \le i \le n$. Thus, from (A.1), we obtain the inclusion relation

$$\begin{aligned} &\left\{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 > \delta\right\} \\ &\subset \{\mathbb{M}_{ni}(\alpha_i, \boldsymbol{\beta}) \le \mathbb{M}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0),\ 1 \le \exists i \le n,\ \exists(\alpha_i, \boldsymbol{\beta}) \notin B_i(\delta)\} \\ &\subset \left\{\max_{1 \le i \le n} \sup_{(\alpha_i, \boldsymbol{\beta}) \in B_i(\delta)} |\Delta_{ni}(\alpha_i, \boldsymbol{\beta}) - \mathrm{E}[\Delta_{ni}(\alpha_i, \boldsymbol{\beta})]| \ge \epsilon_\delta\right\}. \end{aligned}$$

Therefore, it suffices to show that for every $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}\left\{\max_{1 \le i \le n} \sup_{(\alpha_i, \boldsymbol{\beta}) \in B_i(\delta)} |\Delta_{ni}(\alpha_i, \boldsymbol{\beta}) - \mathrm{E}[\Delta_{ni}(\alpha_i, \boldsymbol{\beta})]| > \epsilon\right\} = 0. \quad \text{(A.2)}$$

Because of the subadditivity of the probability measure, it suffices to prove that for every $\epsilon > 0$,

$$\max_{1\le i\le n} \mathrm{P}\left\{ \sup_{(\alpha,\boldsymbol{\beta})\in B_i(\delta)} |\Delta_{ni}(\alpha,\boldsymbol{\beta}) - \mathrm{E}[\Delta_{ni}(\alpha,\boldsymbol{\beta})]| > \epsilon \right\} = o(n^{-1}). \tag{A.3}$$

We follow the proof of Lemma 7 in Fernandez-Val (2005) to show (A.3). Without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. Then, $B_i(\delta)$ is independent of i and write $B_i(\delta) = B(\delta)$ for simplicity. Put $g_{\alpha,\boldsymbol{\beta}}(u,\boldsymbol{x}) := \rho_\tau(u - \alpha - \boldsymbol{x}'\boldsymbol{\beta}) - \rho_\tau(u)$. Observe that $|g_{\alpha,\boldsymbol{\beta}}(u,\boldsymbol{x}) - g_{\bar{\alpha},\bar{\boldsymbol{\beta}}}(u,\boldsymbol{x})| \le C(1 + \|\boldsymbol{x}\|_1)(|\alpha - \bar{\alpha}| + \|\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}\|_1)$ for some universal constant $C > 0$. Put $L(\boldsymbol{x}) := C(1 + \|\boldsymbol{x}\|)$ and $\kappa := \sup_{i\ge 1} \mathrm{E}[L(\boldsymbol{x}_{i1})]$. Since $B(\delta)$ is a compact subset of $\mathbb{R}^{p+1}$, there exist K ℓ_1-balls with centers $(\alpha^{(j)}, \boldsymbol{\beta}^{(j)})$, $j = 1,\dots,K$ and radius $\epsilon/(7\kappa)$ such that the collection of these balls covers $B(\delta)$. Note that K is independent of i and can be chosen such that $K = K(\epsilon) = O(\epsilon^{-p-1})$ as $\epsilon \to 0$. Now, for each $(\alpha,\boldsymbol{\beta}) \in B(\delta)$, there is a $j \in \{1,\dots,K\}$ such that $|g_{\alpha,\boldsymbol{\beta}}(u,\boldsymbol{x}) - g_{\alpha^{(j)},\boldsymbol{\beta}^{(j)}}(u,\boldsymbol{x})| \le L(\boldsymbol{x})\epsilon/(7\kappa)$, which leads to $|\Delta_{ni}(\alpha,\boldsymbol{\beta}) - \mathrm{E}[\Delta_{ni}(\alpha,\boldsymbol{\beta})]| \le |\Delta_{ni}(\alpha^{(j)},\boldsymbol{\beta}^{(j)}) - \mathrm{E}[\Delta_{ni}(\alpha^{(j)},\boldsymbol{\beta}^{(j)})]| + \{\epsilon/(7\kappa)\} \cdot |T^{-1}\sum_{t=1}^T \{L(\boldsymbol{x}_{it}) - \mathrm{E}[L(\boldsymbol{x}_{i1})]\}| + 2\epsilon/7$. Therefore, we have

$$\begin{aligned}
&\mathrm{P}\left\{ \sup_{(\alpha,\boldsymbol{\beta})\in B(\delta)} |\Delta_{ni}(\alpha,\boldsymbol{\beta}) - \mathrm{E}[\Delta_{ni}(\alpha,\boldsymbol{\beta})]| > \epsilon \right\} \\
&\quad \le \sum_{j=1}^K \mathrm{P}\left\{ |\Delta_{ni}(\alpha^{(j)},\boldsymbol{\beta}^{(j)}) - \mathrm{E}[\Delta_{ni}(\alpha^{(j)},\boldsymbol{\beta}^{(j)})]| > \frac{\epsilon}{3} \right\} \\
&\quad\quad + \mathrm{P}\left\{ \frac{1}{T}\left| \sum_{t=1}^T \{L(\boldsymbol{x}_{it}) - \mathrm{E}[L(\boldsymbol{x}_{i1})]\} \right| > \frac{7\kappa}{3} \right\}.
\end{aligned} \tag{A.4}$$

Since $\sup_{i\ge1} \mathrm{E}[L^{2s}(\boldsymbol{x}_{i1})] < \infty$, application of the Marcinkiewicz-Zygmung inequality (see Corollary 2 in Chow and Teicher, 1997, p. 387) implies that both terms in the right hand side of (A.4) are $O(T^{-s})$ uniformly over $1 \le i \le n$. Because of the hypothesis on T, they are $o(n^{-1})$, leading to (A.3).

Step 2. Next, we shall show that $\max_{1\le i\le n} |\hat{\alpha}_i - \alpha_{i0}| \xrightarrow{p} 0$. Recall that $\hat{\alpha}_i = \arg\min_\alpha \mathbb{M}_{ni}(\alpha, \hat{\boldsymbol{\beta}})$. Fix any $\delta > 0$. For each $\alpha_i \in \mathbb{R}$ such that $|\alpha_i - \alpha_{i0}| > \delta$, define $\tilde{\alpha}_i = r_i\alpha_i + (1 - r_i)\alpha_{i0}$, where $r_i = \delta/|\alpha_i - \alpha_{i0}|$. Because of the convexity of the criterion function, we have

$$\begin{aligned}
r_i\{&\mathbb{M}_{ni}(\alpha_i,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0},\hat{\boldsymbol{\beta}})\} \ge \mathbb{M}_{ni}(\tilde{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0},\hat{\boldsymbol{\beta}}) \\
&= \mathbb{M}_{ni}(\tilde{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0},\boldsymbol{\beta}_0) - \{\mathbb{M}_{ni}(\alpha_{i0},\hat{\boldsymbol{\beta}}) - \mathbb{M}_{ni}(\alpha_{i0},\boldsymbol{\beta}_0)\} \\
&= \{\Delta_{ni}(\tilde{\alpha}_i,\hat{\boldsymbol{\beta}}) - \mathrm{E}[\Delta_{ni}(\tilde{\alpha}_i,\boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} - \{\Delta_{ni}(\alpha_{0i},\hat{\boldsymbol{\beta}}) - \mathrm{E}[\Delta_{ni}(\alpha_{0i},\boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} \\
&\quad + \mathrm{E}[\Delta_{ni}(\tilde{\alpha}_i,\boldsymbol{\beta}_0)] + \{\mathrm{E}[\Delta_{ni}(\tilde{\alpha}_i,\boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \mathrm{E}[\Delta_{ni}(\tilde{\alpha}_i,\boldsymbol{\beta}_0)]\} + \mathrm{E}[\Delta_{ni}(\alpha_{0i},\boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}.
\end{aligned}$$

It is seen from condition (A3) that the third term in the right hand side is greater than or equal to ϵ_δ. Thus, we obtain the inclusion relation

$$\begin{aligned}
&\{|\hat{\alpha}_i - \alpha_{i0}| > \delta,\ 1 \le \exists i \le n\} \\
&\subset \{\mathbb{M}_{ni}(\alpha_i, \hat{\boldsymbol{\beta}}) \le \mathbb{M}_{ni}(\alpha_{i0}, \hat{\boldsymbol{\beta}}),\ 1 \le \exists i \le n,\ \exists \alpha_i \in \mathbb{R} \text{ s.t. } |\alpha_i - \alpha_{i0}| > \delta\} \\
&\subset \left\{ \max_{1 \le i \le n} \sup_{|\alpha - \alpha_{i0}| \le \delta} |\Delta_{ni}(\alpha, \hat{\boldsymbol{\beta}}) - \mathrm{E}[\Delta_{ni}(\alpha, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}| \ge \frac{\epsilon_\delta}{4} \right\} \\
&\quad \cup \left\{ \max_{1 \le i \le n} \sup_{|\alpha - \alpha_{i0}| \le \delta} ||\mathrm{E}[\Delta_{ni}(\alpha, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \mathrm{E}[\Delta_{ni}(\alpha, \boldsymbol{\beta}_0)]| \ge \frac{\epsilon_\delta}{4} \right\} \\
&=: A_{1n} \cup A_{2n}.
\end{aligned}$$

Since $\hat{\boldsymbol{\beta}}$ is consistent by Step 1, and especially $\hat{\boldsymbol{\beta}} = O_p(1)$, from (A.2), it is shown that $\mathrm{P}(A_{1n}) \to 0$. Finally, since

$$|\mathrm{E}[\Delta_{ni}(\alpha, \boldsymbol{\beta})] - \mathrm{E}[\Delta_{ni}(\alpha, \boldsymbol{\beta}_0)]| \le 2\mathrm{E}[\|\boldsymbol{x}_{i1}\|]\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|,$$

and $\sup_{i \ge 1} \mathrm{E}[\|\boldsymbol{x}_{i1}\|] \le 1 + \sup_{i \ge 1} \mathrm{E}[\|\boldsymbol{x}_{i1}\|^{2s}] < \infty$, consistency of $\hat{\boldsymbol{\beta}}$ implies that $\mathrm{P}(A_{2n}) \to 0$. Therefore, we complete the proof. $\square$

Remark A.1. If $\sup_{i \ge 1} \|\boldsymbol{x}_{i1}\| \le M$ (a.s.) for some constant M, we may take $L(\boldsymbol{x}) \equiv C(1+M)$ and the second term in the right hand side of (A.4) will vanish. In this case, we can apply Hoeffding's inequality to the first term in the right hand side of (A.4) and the probability in (A.3) is bounded by $D\exp(-DT)$ for some positive constant D that depends on ϵ but not on i. Therefore, the conclusion of Theorem 2.1 holds when $\log n/T \to 0$ as $n \to \infty$ in this case.

A.2 Proof of Theorem 2.2

Define

$$\begin{aligned}
\mathbb{H}_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}) &:= \frac{1}{T}\sum_{t=1}^{T}\{\tau - I(y_{it} \le \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})\}, \\
H_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta}) &:= \mathrm{E}[\mathbb{H}_{ni}^{(1)}(\alpha_i, \boldsymbol{\beta})] = \mathrm{E}[\{\tau - F_i(\alpha_i - \alpha_{i0} + \boldsymbol{x}_{i1}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_{i1})\}], \\
\mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &:= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\tau - I(y_{it} \le \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})\}\boldsymbol{x}_{it}, \\
H_{n}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &:= \mathrm{E}[\mathbb{H}_{n}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta})] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\{\tau - F_i(\alpha_i - \alpha_{i0} + \boldsymbol{x}_{i1}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_{i1})\}\boldsymbol{x}_{i1}].
\end{aligned}$$

Note that $\mathbb{H}_{ni}(\alpha_i, \boldsymbol{\beta})$ depends on n since T does. The $(n+p)$ dimensional vector of functions

$$[\mathbb{H}_{n1}^{(1)}(\alpha_1, \boldsymbol{\beta}), \ldots, \mathbb{H}_{nn}^{(1)}(\alpha_n, \boldsymbol{\beta}), \mathbb{H}_{n}^{(2)\prime}(\boldsymbol{\alpha}, \boldsymbol{\beta})]'$$

are called the scores for problem (2.2).

Before starting the proof, we introduce some notation used in empirical process theory. Let $\mathcal{F}$ be a class of measurable functions on a measurable space $(S, \mathcal{S})$. For a process $Z(f)$ defined on $\mathcal{F}$, $\|Z(f)\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |Z(f)|$. For a probability measure Q on $(S, \mathcal{S})$ and $\epsilon > 0$, let $N(\mathcal{F}, L_2(Q), \epsilon)$ denote the ϵ-covering number of $\mathcal{F}$ with respect to the $L_2(Q)$ norm $\|\cdot\|_{L_2(Q)}$. For the definition of a Vapnik-Červonenkis (VC) subgraph class, we refer to van der Vaart and Wellner (1996), Section 2.6.

Proof of Theorem 2.2. We divide the proof into four steps.

Step 1 (Asymptotic representation). We shall show that

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) = {} & \Gamma_n^{-1}\{-n^{-1}\textstyle\sum_{i=1}^n \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} \\ & - \Gamma_n^{-1}[n^{-1}\textstyle\sum_{i=1}^n \boldsymbol{\gamma}_i\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\}] \\ & + \Gamma_n^{-1}\{\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} \\ & + O_p\{T^{-1} + \max_{1 \le i \le n} |\hat{\alpha}_i - \alpha_{i0}|^2\}. \end{aligned} \tag{A.5}$$

Because of the computational property of the QR estimator (see equation (3.10) in Gutenbrunner and Jureckova, 1992), it is shown that $\max_{1 \le i \le n} |\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}})| = O_p(T^{-1})$. Thus, uniformly over $1 \le i \le n$, we have

$$O_p(T^{-1}) = \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) + H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) + \{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\}.$$

Expanding $H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}})$ around $(\alpha_{i0}, \boldsymbol{\beta}_0)$, we have

$$\begin{aligned}\hat{\alpha}_i - \alpha_{i0} = {} & \{f_i(0)\}^{-1}\mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0) - \boldsymbol{\gamma}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ & + \{f_i(0)\}^{-1}\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\} + r_{ni}, \end{aligned} \tag{A.6}$$

where $\max_{1 \le i \le n} |r_{ni}| = o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) + O_p\{T^{-1} + \max_{1 \le i \le n} |\hat{\alpha}_i - \alpha_{i0}|^2\}$.

Similarly, the computational property of the QR estimator implies that $\|\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})\| = O_p\{T^{-1} \max_{1 \le i \le n, 1 \le t \le T} \|\boldsymbol{x}_{it}\|\} = O_p(T^{-1})$, from which we have

$$O_p(T^{-1}) = \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) + \{\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\}. \tag{A.7}$$

Use Taylor's theorem to obtain

$$\begin{aligned} H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = {} & -\left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}](\hat{\alpha}_i - \alpha_{i0})\right\} \\ - & \left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}']\right\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) + O_p\left\{\max_{1 \le i \le n} |\hat{\alpha}_i - \alpha_{i0}|^2\right\}. \end{aligned} \tag{A.8}$$

Plugging (A.6) into (A.8) leads to

$$\begin{aligned} H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) &= -\frac{1}{n}\sum_{i=1}^{n} \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\boldsymbol{\gamma}_i - \Gamma_n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\gamma}_i\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\} \\ &\quad + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) + O_p\left\{T^{-1} + \max_{1\le i\le n}|\hat{\alpha}_i - \alpha_{i0}|^2\right\}. \end{aligned} \tag{A.9}$$

Combining (A.7) and (A.9) yields the desired representation. The next two steps are devoted to determining the order of the remainder terms in (A.5).

Step 2 (Stochastic equicontinuity). Take $\delta_n \to 0$ such that $\max_{1\le i\le n}|\hat{\alpha}_i - \alpha_{i0}| \vee \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(\delta_n)$.[4] We shall show that

$$\|n^{-1}\textstyle\sum_{i=1}^{n}\boldsymbol{\gamma}_i\{\mathbb{H}_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\}\| = O_p(d_n), \tag{A.10}$$

$$\|\mathbb{H}_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_n^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\| = O_p(d_n). \tag{A.11}$$

where $d_n := T^{-1}|\log \delta_n| + T^{-1/2}\delta_n^{1/2}|\log \delta_n|^{1/2}$.

We only prove (A.10) since the proof of (A.11) is analogous.[5] Without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$. Put $g_{\alpha,\boldsymbol{\beta}}(u, \boldsymbol{x}) := I(u \le \alpha + \boldsymbol{x}'\boldsymbol{\beta}) - I(u \le 0)$, $\mathcal{G}_\delta := \{g_{\alpha,\boldsymbol{\beta}} : |\alpha| \le \delta, \|\boldsymbol{\beta}\| \le \delta\}$ and $\xi_{it} := (u_{it}, \boldsymbol{x}_{it})$. Since $\boldsymbol{\gamma}_i$ is bounded over i, it suffices to show that

$$\max_{1\le i\le n} \mathrm{E}\left[\left\|\sum_{t=1}^{T}\{g(\xi_{it}) - \mathrm{E}[g(\xi_{i1})]\}\right\|_{\mathcal{G}_{\delta_n}}\right] = O(d_n T).$$

We make use of Proposition B.1 in Appendix B to show this. For each i, let $\epsilon_{i1}, \ldots, \epsilon_{iT}$ be i.i.d. random variables with $\mathrm{P}(\epsilon_{it} = \pm 1) = 1/2$ and independent of $\{\xi_{it}, t = 1, \ldots, T\}$. Observe from Lemma 2.3.1 in van der Vaart and Wellner (1996) that

$$\mathrm{E}\left[\left\|\sum_{t=1}^{T}\{g(\xi_{it}) - \mathrm{E}[g(\xi_{i1})]\}\right\|_{\mathcal{G}_{\delta_n}}\right] \le 2\mathrm{E}\left[\left\|\sum_{t=1}^{T}\epsilon_{it}\{g(\xi_{it}) - \mathrm{E}[g(\xi_{i1})]\}\right\|_{\mathcal{G}_{\delta_n}}\right]. \tag{A.12}$$

We apply Proposition B.1 to $\tilde{\mathcal{G}}_{i,\delta_n} := \{g - \mathrm{E}[g(\xi_{i1})] : g \in \mathcal{G}_{\delta_n}\}$. Observe that $\tilde{\mathcal{G}}_{i,\delta_n}$ is pointwise measurable and each element of $\tilde{\mathcal{G}}_{i,\delta_n}$ is bounded by 4. Because of Lemmas 2.6.15 and 2.6.18 of van der Vaart and Wellner (1996), $\mathcal{G}_\infty = \{g_{\alpha,\boldsymbol{\beta}} : \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^p\}$ is a VC subgraph class. Thus, from Theorem 2.6.7 of van der Vaart and Wellner (1996), the fact that $\tilde{\mathcal{G}}_{i,\delta_n} \subset \{g - \mathrm{E}[g(\xi_{i1})] : g \in \mathcal{G}_\infty\}$, and a simple estimate of covering numbers, there exist

[4] For $a, b \in \mathbb{R}$, $a \vee b := \max\{a, b\}$.

[5] Though the present proof requires $\boldsymbol{x}_{i1}$ to be bounded, it is possible to use Theorem 2.14.1 of van der Vaart and Wellner (1996) to show (A.11), which only requires that $\sup_{i\ge 1}\mathrm{E}[\|\boldsymbol{x}_{i1}\|^2] < \infty$. However, recall that condition (B1) is used to ensure (A.7).

constants $A \geq 3\sqrt{e}$ and $v \geq 1$ independent of i and n such that $N(\tilde{\mathcal{G}}_{i,\delta_n}, L_2(Q), 4\epsilon) \leq (A/\epsilon)^v$ for $0 < \epsilon < 1$ and for every probability measure Q on $\mathbb{R}^{p+1}$. Observe that $\mathrm{E}[g^2_{\alpha,\boldsymbol{\beta}}(\xi_{i1})] = \mathrm{E}[|F_i(\alpha + \boldsymbol{x}'_{i1}\boldsymbol{\beta}|\boldsymbol{x}_{i1}) - F_i(0|\boldsymbol{x}_{i1})|] \leq A_0(|\alpha| + M\|\boldsymbol{\beta}\|)$ and take $D := A_0(1+M)$. Then, we can see that $\tilde{\mathcal{G}}_{i,\delta_n}$ satisfies all the conditions of Proposition B.1 with $F \equiv 4$, $U = 4$ and $\sigma^2 = D\delta_n$, and the constants A, v and D are independent of i and n, which implies that the right hand side of (A.12) is $O(d_n T)$ uniformly over $1 \leq i \leq n$.

Step 3 (Convergence rates). We shall show that

$$\max_{1\leq i\leq n} |\hat{\alpha}_i - \alpha_{i0}| = O_p\{(T/\log n)^{-1/2}\},\ \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = o_p(T^{-1/2}).$$

Because of consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and the result given in Step 2, the second and third terms in the right hand side of equation (A.5) is $o_p(T^{-1/2})$, which implies that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\{\max_{1\leq i\leq n} |\hat{\alpha}_i - \alpha_{i0}|^2\} + o_p(T^{-1/2})$. Thus, from (A.6), $\max_{1\leq i\leq n} |\hat{\alpha}_i - \alpha_{i0}|$ is bounded by

$$\text{const.}\times\left\{\max_{1\leq i\leq n} |\mathbb{H}^{(1)}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)| + \max_{1\leq i\leq n} |\mathbb{H}^{(1)}_{ni}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H^{(1)}_{ni}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}^{(1)}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)|\right\} + o_p(T^{-1/2}),$$

with probability approaching one. First, observe that for $K > 0$,

$$\mathrm{P}\left\{\max_{1\leq i\leq n} |\mathbb{H}^{(1)}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)| > (T/\log n)^{-1/2}K\right\} \leq \sum_{i=1}^{n} \mathrm{P}\left\{|\mathbb{H}^{(1)}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)| > (T/\log n)^{-1/2}K\right\},$$

and the right hand side is bounded by $2n^{1-K^2/2}$ by Hoeffding's inequality. This implies that $\max_{1\leq i\leq n} |\mathbb{H}^{(1)}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)| = O_p\{(T/\log n)^{-1/2}\}$.

We next show that

$$\max_{1\leq i\leq n} |\mathbb{H}^{(1)}_{ni}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H^{(1)}_{ni}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}^{(1)}_{ni}(\alpha_{i0}, \boldsymbol{\beta}_0)| = o_p\{(T/\log n)^{-1/2}\},$$

which will lead to the desired result. Again, without loss of generality, we may assume that $\alpha_{i0} = 0$ and $\boldsymbol{\beta}_0 = 0$. Let $g_{\alpha,\boldsymbol{\beta}}, \mathcal{G}_\delta, \tilde{\mathcal{G}}_{i,\delta}$ and ξ_{it} be the same as those given in Step 2. Because of consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and the subadditivity of the probability measure, it suffices to show that for every $\epsilon > 0$, there exists a sufficiently small $\delta > 0$ such that

$$\max_{1\leq i\leq n} \mathrm{P}\left\{\left\|\sum_{t=1}^{T}\{g(\xi_{it}) - \mathrm{E}[g(\xi_{i1})]\}\right\|_{\mathcal{G}_\delta} > (T\log n)^{1/2}\epsilon\right\} = o(n^{-1}).$$

To show this, we make use of Proposition B.2 in Appendix B. Fix $\epsilon > 0$. From Step 2, observe that $\tilde{\mathcal{G}}_{i,\delta}$ satisfies the conditions of Proposition B.1 with $F \equiv 4$, $U = 4$, $\sigma^2 = D\delta$, where $D := A_0(1+M)$, and the constants A and v are independent of i and δ. Let L, C be the constants given in Proposition B.2. Note that L and C depend only on A and v. Take $\delta > 0$ such that $\delta < \min\{4/D, \epsilon^2\varphi(L,C)/D\}$ and fix it, where, as in Proposition

B.2, $\varphi(L,C) := \log(1 + C/(4L))/(LC)$. Then, from Proposition B.2 and the fact that $\log n/T \to 0$, there exists a positive integer n_0 such that for $n \geq n_0$,

$$\max_{1\leq i\leq n} \mathrm{P}\left\{\left\|\sum_{t=1}^{T}\{g(\xi_{it}) - \mathrm{E}[g(\xi_{i1})]\}\right\|_{\mathcal{G}_\delta} > (T\log n)^{1/2}\epsilon\right\} \leq L\exp\left\{-\frac{\epsilon^2\varphi(L,C)\log n}{D\delta}\right\},$$

which is $o(n^{-1})$ because of the choice of δ.

Step 4 (Conclusion). From Step 3, we may take $\delta_n = (T/\log n)^{-1/2}$ in Step 2. Thus, from Step 1, we obtain

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + o_p(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|) = \Gamma_n^{-1}\{-n^{-1}\textstyle\sum_{i=1}^{n}\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\}\\ + O_p\{T^{-3/4}(\log n)^{3/4}\}, \quad \text{(A.13)}\end{aligned}$$

which leads to (2.4). Finally, if $n^2(\log n)^3/T \to 0$, $O_p\{T^{-3/4}(\log n)^{3/4}\} = o_p\{(nT)^{-1/2}\}$ and $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p\{(nT)^{-1/2}\}$. Application of the Lyapunov central limit theorem shows that

$$\sqrt{nT}\{-n^{-1}\textstyle\sum_{i=1}^{n}\mathbb{H}_{ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_n^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\} \xrightarrow{d} N\{\mathbf{0},\tau(1-\tau)V\}.$$

Therefore, we obtain (2.5). □

Remark A.2. The reason why the order of the remainder term in (A.13) is $O_p\{T^{-3/4}(\log n)^{3/4}\}$ and not $O_p(T^{-1})$ is that the exponent of δ_n inside the O_p terms in the right hand side of equations (A.10) and (A.11) is 1/2 and not 1. Recall the definition of $g_{\alpha,\boldsymbol{\beta}}$ given in Step 2. Since $g_{\alpha,\boldsymbol{\beta}}$ is not differentiable with respect to $(\alpha,\boldsymbol{\beta})$, $\mathrm{E}[g_{\alpha,\boldsymbol{\beta}}^2(\xi_{i1})]$ is bounded by const. $\times$ $(|\alpha| + \|\boldsymbol{\beta}\|)$ but not by const. $\times$ $(|\alpha|^2 + \|\boldsymbol{\beta}\|^2)$, which results in the exponent 1/2 of δ_n. Note that if $g_{\alpha,\boldsymbol{\beta}}$ were smooth in $(\alpha,\boldsymbol{\beta})$, we could use Taylor's theorem to bound $\mathrm{E}[g_{\alpha,\boldsymbol{\beta}}^2(\xi_{i1})]$ by const. $\times$ $(|\alpha|^2 + \|\boldsymbol{\beta}\|^2)$. In that case, the exponent of δ_n would be 1, leading to the $O_p(T^{-1})$ rate of the remainder terms (we have ignored the $\log n$ term).

B Inequalities from empirical process theory

In this appendix, we introduce two inequalities from empirical process theory that were used in the proof of Theorem 2.2. Let $\xi_1,\dots,\xi_n$ be i.i.d. random variables taking values in a measurable space $(S,\mathcal{S})$ and let $\epsilon_1,\dots,\epsilon_n$ be i.i.d. random variables independent of $\{\xi_i, i = 1,\dots,n\}$ with $\mathrm{P}(\epsilon_i = \pm 1) = 1/2$. For a function f on S, let $\|f\|_\infty := \sup_{x\in S}|f(x)|$. The next proposition is a moment inequality of symmetrized empirical processes, which is due to Proposition 2.2 in Gine and Guillou (2001). The same inequality for indicator functions is found in Talagrand (1994), Proposition 6.2. To avoid the measurability problem, we assume $\mathcal{F}$ to be a pointwise measurable class of functions, i.e., each element of $\mathcal{F}$ is measurable and there exists a countable subset $\mathcal{G} \subset \mathcal{F}$ such that for each $f \in \mathcal{F}$, there exists a sequence $\{g_m\} \subset \mathcal{G}$ with $g_m(\xi) \to f(\xi)$ for all $\xi \in S$. This condition is discussed in Section 2.3 of van der Vaart and Wellner (1996).

Proposition B.1. *Let $\mathcal{F}$ be a uniformly bounded, pointwise measurable class of functions on $(S, \mathcal{S})$ such that for some constants $A \geq 3\sqrt{e}$ and $v \geq 1$, $N(\mathcal{F}, L_2(Q), \epsilon\|F\|_{L_2(Q)}) \leq (A/\epsilon)^v$ for $0 < \epsilon < 1$ and for every probability measure Q on $(S, \mathcal{S})$ with $\|F\|_{L_2(Q)} > 0$, where $F(x) \geq \sup_{f\in\mathcal{F}} |f(x)|$ for all $x \in S$. Let $\sigma^2 \geq \sup_{f\in\mathcal{F}} \mathrm{E}[f^2(\xi_1)]$ and $U \geq \|F\|_\infty$ be such that $0 < \sigma \leq U$. Then, for all $n \geq 1$,*

$$\mathrm{E}\left[\left\|\sum_{i=1}^n \epsilon_i f(\xi_i)\right\|_{\mathcal{F}}\right] \leq C\left[vU\log\frac{AU}{\sigma} + \sqrt{v}\sqrt{n}\sigma\sqrt{\log\frac{AU}{\sigma}}\right],$$

where C is a universal constant.

The next proposition is an exponential inequality for centered empirical processes, which originates from Talagrand (1994, 1996). The current form of the inequality is due to Corollary 2.2 in Gine and Guillou (2002). Proposition B.2 is derived from combining Talagrand's (1996) Theorem 1.4 and the moment inequality in Proposition B.1.[6] See the derivation of Proposition 2.2 in Gine and Guillou (2001).

Proposition B.2. *Suppose that $\mathcal{F}$ satisfies the conditions of Proposition B.1 and $\mathrm{E}[f(\xi_1)] = 0$ for all $f \in \mathcal{F}$. If, moreover, $0 < \sigma < U/2$ and $\sqrt{n}\sigma \geq U\sqrt{\log(U/\sigma)}$, there exits positive constants L and C depending only on A and v such that for all t satisfying*

$$C\sqrt{n}\sigma\sqrt{\log\frac{U}{\sigma}} \leq t \leq C\frac{n\sigma^2}{U},$$

we have

$$\mathrm{P}\left\{\left\|\sum_{i=1}^n f(\xi_i)\right\|_{\mathcal{F}} > t\right\} \leq L\exp\left\{-\frac{\varphi(L,C)t^2}{n\sigma^2}\right\},$$

where $\varphi(L, C) := \log(1 + C/(4L))/(LC)$.

References

Abrevaya, J., and C. M. Dahl (2008): "The Effects of Birth Inputs on Birthweight: Evidence From Quantile Estimation on Panel Data," *Journal of Business and Economic Statistics*, 26, 379–397.

[6] Talagrand's (1996) Theorem 1.4 assumes $\mathcal{F}$ to be a countable class. Clearly, this condition can be weakened to the case where $\mathcal{F}$ is pointwise measurable. Massart (2000) provided a simple proof of this theorem.

Angrist, J., V. Chernozhukov, and I. Fernandez-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Arcones, M. A. (1998): "Second Order Representations of the Least Absolute Deviation Regression Estimator," *Annals of the Institute of Statistical Mathematics*, 50, 87–117.

Buchinsky, M. (1995): "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics*, 68, 303–338.

Cameron, A. C., and P. K. Trivedi (2005): *Microeconometrics: Methods and Applications.* Cambridge University Press, Cambridge, Massachusetts.

Canay, I. A. (2008): "A Note on Quantile Regression for Panel Data Models," Mimeo, Northwestern University.

Chernozhukov, V., I. Fernandez-Val, and W. Newey (2009): "Quantile and Average Effects in Nonseparable Panel Models," CEMMAP working paper CWP29/09.

Chow, Y. S., and H. Teicher (1997): *Probability Theory (3rd Edition).* Springer-Verlag, New York.

Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference Approach," *Annals of Statistics*, 2, 267–277.

Fernandez-Val, I. (2005): "Bias Correction in Panel Data Models with Individual Specific Parameters," Mimeo.

Gine, E., and A. Guillou (2001): "On Consistency of Kernel Density Estimators for Randomly Censored Data: Rates Holding Uniformly over Adaptive Intervals," *Annales de l'Institut Henri Poincare*, 37, 503–522.

——— (2002): "Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators," *Annales de l'Institut Henri Poincare*, 38, 907–921.

Goncalves, S. (2008): "The Moving Blocks Bootstrap for Panel Linear Regression Models with Individual Fixed Effects," Mimeo.

Graham, B. S., J. Hahn, and J. L. Powell (2009): "The Incidental Parameter Problem in a Non-Differentiable Panel Data Model," *Economics Letters*, 105, 181–182.

Gutenbrunner, C., and J. Jureckova (1992): "Regression Rank Scores and Regression Quantiles," *Annals of Statistics*, 20, 305–330.

Hahn, J. (1995): "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11, 105–121.

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

He, X., and Q.-M. Shao (2000): "On Parameters of Increasing Dimensions," *Journal of Multivariate Analysis*, 73, 120–135.

Horowitz, J. L. (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66, 1327–1351.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equation Models without Additivity," *Econometrica*, 77, 1481–1512.

Kapetanios, G. (2008): "A Bootstrap Procedure for Panel Datasets with Many Cross-Sectional Units," *Econometrics Journal*, 11, 377–395.

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

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

Koenker, R., and G. W. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–49.

Koenker, R., and K. Hallock (2000): "Quantile Regression: An Introduction," Manuscript, University of Illinois at Urbana-Champaign.

Lamarche, C. (2006): "Robust Penalized Quantile Regression Estimation for Panel Data," Mimeo, University of Oklahoma.

Lancaster, T. (2000): "The Incidental Parameter Problem since 1948," *Journal of Econometrics*, 95, 391–413.

Li, H., B. G. Lindsay, and R. P. Waterman (2003): "Efficiency of Projected Score Methods in Rectangular Array Asymptotics," *Journal of the Royal Statistical Society, Series B*, 65, 191–208.

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

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

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Rosen, A. (2009): "Set Identification via Quantile Restrictions in Short Panels," CEMMAP working paper CWP26/09.

Ruppert, D., and R. J. Caroll (1980): "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828–838.

Talagrand, M. (1994): "Sharper Bounds for Gaussian and Empirical Processes," *Annals of Probability*, 22, 28–76.

——— (1996): "New Concentration Inequalities in Product Spaces," *Inventiones Mathematicae*, 126, 505–563.

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

Table 1: Bias and standard deviation of $\hat{\beta}(\tau)$. Location shift model.

		$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$			$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$		
	n/T	5	10	50	5	10	50
$\tau = 0.25$	25	0.002	-0.000	0.000	0.004	0.004	0.001
		[0.056]	[0.037]	[0.016]	[0.074]	[0.048]	[0.021]
	50	0.000	-0.000	0.000	0.003	0.002	0.000
		[0.037]	[0.026]	[0.011]	[0.052]	[0.033]	[0.015]
$\tau = 0.50$	25	0.000	-0.001	0.000	0.005	0.004	0.001
		[0.050]	[0.035]	[0.015]	[0.099]	[0.072]	[0.031]
	50	0.000	0.000	0.000	0.002	0.002	0.000
		[0.035]	[0.025]	[0.011]	[0.069]	[0.050]	[0.022]
$\tau = 0.75$	25	-0.001	-0.001	0.000	0.003	0.004	0.001
		[0.056]	[0.037]	[0.016]	[0.153]	[0.106]	[0.047]
	50	-0.001	0.000	0.000	0.001	0.002	0.000
		[0.039]	[0.026]	[0.011]	[0.104]	[0.075]	[0.034]

Notes: Standard deviation in brackets. The bias is computed as $\mathrm{E}[\hat{\beta}(\tau) - \beta_0(\tau)]$.

Table 2: Bias and standard deviation of $\hat{\beta}(\tau)$. Location-scale shift model.

		$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$			$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$		
	n/T	5	10	50	5	10	50
$\tau = 0.25$	25	0.028	0.013	0.002	0.021	0.008	0.001
		[0.118]	[0.080]	[0.035]	[0.159]	[0.105]	[0.046]
	50	0.027	0.012	0.002	0.013	0.004	-0.000
		[0.083]	[0.056]	[0.025]	[0.114]	[0.074]	[0.032]
$\tau = 0.50$	25	0.000	0.001	-0.000	-0.018	-0.008	-0.002
		[0.107]	[0.075]	[0.032]	[0.214]	[0.156]	[0.068]
	50	0.001	-0.000	-0.000	-0.029	-0.011	-0.003
		[0.076]	[0.052]	[0.023]	[0.153]	[0.108]	[0.047]
$\tau = 0.75$	25	-0.027	-0.013	-0.003	-0.102	-0.052	-0.010
		[0.118]	[0.080]	[0.035]	[0.328]	[0.233]	[0.104]
	50	-0.026	-0.014	-0.003	-0.111	-0.056	-0.011
		[0.084]	[0.056]	[0.025]	[0.231]	[0.164]	[0.074]

Notes: Standard deviation in brackets. The bias is computed as $\mathrm{E}[\hat{\beta}(\tau) - \beta_0(\tau)]$.

Table 3: Rejection rate and length. Location shift model.

		Cross-sectional			Temporal			Cross-S.&Temp.		
$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$										
	n / T	5	10	50	5	10	50	5	10	50
$\tau = 0.25$	25	0.128	0.144	0.148	0.098	0.054	0.072	0.016	0.008	0.020
		[0.182]	[0.116]	[0.049]	[0.183]	[0.127]	[0.055]	[0.250]	[0.165]	[0.071]
	50	0.100	0.170	0.110	0.072	0.088	0.066	0.022	0.008	0.012
		[0.127]	[0.081]	[0.036]	[0.125]	[0.089]	[0.039]	[0.172]	[0.116]	[0.051]
$\tau = 0.50$	25	0.116	0.138	0.130	0.074	0.050	0.066	0.004	0.018	0.012
		[0.160]	[0.111]	[0.046]	[0.181]	[0.126]	[0.051]	[0.244]	[0.162]	[0.067]
	50	0.090	0.132	0.136	0.058	0.054	0.052	0.022	0.006	0.010
		[0.113]	[0.078]	[0.033]	[0.124]	[0.089]	[0.036]	[0.168]	[0.113]	[0.048]
$\tau = 0.75$	25	0.158	0.126	0.112	0.084	0.080	0.060	0.012	0.016	0.014
		[0.180]	[0.116]	[0.050]	[0.182]	[0.127]	[0.055]	[0.249]	[0.166]	[0.072]
	50	0.098	0.132	0.104	0.072	0.058	0.042	0.028	0.018	0.016
		[0.126]	[0.082]	[0.036]	[0.126]	[0.090]	[0.039]	[0.174]	[0.118]	[0.051]
$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$										
$\tau = 0.25$	25	0.162	0.116	0.120	0.076	0.088	0.056	0.020	0.014	0.016
		[0.244]	[0.156]	[0.066]	[0.286]	[0.171]	[0.072]	[0.391]	[0.231]	[0.095]
	50	0.104	0.134	0.114	0.060	0.066	0.078	0.020	0.010	0.006
		[0.167]	[0.107]	[0.047]	[0.192]	[0.119]	[0.050]	[0.266]	[0.156]	[0.067]
$\tau = 0.50$	25	0.094	0.124	0.114	0.062	0.054	0.046	0.014	0.014	0.016
		[0.325]	[0.227]	[0.099]	[0.348]	[0.245]	[0.106]	[0.470]	[0.320]	[0.140]
	50	0.120	0.108	0.126	0.060	0.032	0.060	0.020	0.014	0.014
		[0.224]	[0.162]	[0.070]	[0.236]	[0.172]	[0.075]	[0.317]	[0.223]	[0.100]
$\tau = 0.75$	25	0.110	0.106	0.114	0.082	0.062	0.076	0.014	0.014	0.016
		[0.489]	[0.335]	[0.151]	[0.428]	[0.335]	[0.165]	[0.595]	[0.447]	[0.214]
	50	0.098	0.122	0.126	0.090	0.060	0.092	0.028	0.016	0.018
		[0.347]	[0.233]	[0.109]	[0.294]	[0.234]	[0.116]	[0.403]	[0.309]	[.155]

Notes: Length of the confidence interval in brackets. The nominal size is 0.1.

Table 4: Rejection rate and length. Location-scale shift model.

		Cross-sectional			Temporal			Cross-S.&Temp.		
		$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$								
	n / T	5	10	50	5	10	50	5	10	50
$\tau = 0.25$	25	0.174	0.146	0.144	0.464	0.256	0.128	0.200	0.120	0.032
		[0.185]	[0.119]	[0.048]	[0.167]	[0.119]	[0.051]	[0.244]	[0.168]	[0.069]
	50	0.212	0.198	0.120	0.612	0.376	0.144	0.360	0.184	0.040
		[0.133]	[0.084]	[0.034]	[0.120]	[0.084]	[0.036]	[0.171]	[0.119]	[0.049]
$\tau = 0.50$	25	0.136	0.126	0.120	0.114	0.084	0.094	0.020	0.022	0.018
		[0.163]	[0.108]	[0.044]	[0.173]	[0.119]	[0.047]	[0.241]	[0.161]	[0.064]
	50	0.110	0.118	0.104	0.106	0.090	0.104	0.018	0.020	0.016
		[0.119]	[0.077]	[0.031]	[0.119]	[0.084]	[0.033]	[0.171]	[0.115]	[0.046]
$\tau = 0.75$	25	0.202	0.162	0.142	0.412	0.290	0.138	0.196	0.122	0.028
		[0.182]	[0.117]	[0.048]	[0.172]	[0.121]	[0.051]	[0.242]	[0.168]	[0.070]
	50	0.208	0.222	0.126	0.630	0.360	0.120	0.368	0.156	0.042
		[0.119]	[0.077]	[0.031]	[0.117]	[0.085]	[0.036]	[0.174]	[0.120]	[0.050]
		$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \chi^2_3$								
$\tau = 0.25$	25	0.168	0.122	0.132	0.354	0.210	0.082	0.152	0.092	0.024
		[0.262]	[0.166]	[0.063]	[0.283]	[0.175]	[0.069]	[0.404]	[0.244]	[0.094]
	50	0.186	0.140	0.126	0.458	0.206	0.110	0.236	0.104	0.018
		[0.185]	[0.116]	[0.045]	[0.200]	[0.125]	[0.048]	[0.288]	[0.176]	[0.067]
$\tau = 0.50$	25	0.122	0.144	0.144	0.126	0.106	0.086	0.030	0.036	0.026
		[0.346]	[0.224]	[0.091]	[0.355]	[0.244]	[0.098]	[0.489]	[0.338]	[0.134]
	50	0.120	0.112	0.128	0.146	0.100	0.112	0.052	0.024	0.020
		[0.251]	[0.162]	[0.066]	[0.250]	[0.173]	[0.069]	[0.354]	[0.237]	[0.096]
$\tau = 0.75$	25	0.212	0.160	0.146	0.484	0.364	0.142	0.322	0.152	0.042
		[0.521]	[0.342]	[0.141]	[0.419]	[0.331]	[0.153]	[0.598]	[0.470]	[0.206]
	50	0.272	0.224	0.144	0.724	0.500	0.212	0.480	0.266	0.072
		[0.374]	[0.246]	[0.102]	[0.301]	[0.236]	[0.107]	[0.433]	[0.331]	[0.150]

Notes: Length of the confidence interval in brackets. The nominal size is 0.1.