The Econometrics Journal

Econometrics Journal (2011), volume **14**, pp. 368–386.
doi: 10.1111/j.1368-423X.2011.00349.x

A simple approach to quantile regression for panel data

IVAN A. CANAY[†]

[†]*Department of Economics, Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208, USA.*
E-mail: iacanay@northwestern.edu

First version received: May 2010; final version accepted: April 2011

Summary This paper provides a set of sufficient conditions that point identify a quantile regression model with fixed effects. It also proposes a simple transformation of the data that gets rid of the fixed effects under the assumption that these effects are location shifters. The new estimator is consistent and asymptotically normal as both n and T grow.

Keywords: *Deconvolution*, *Panel data models*, *Quantile regression*, *Two-step estimator*.

1. INTRODUCTION

Panel data models and quantile regression models are both widely used in applied econometrics and popular topics of research in theoretical papers. Quantile regression models allow the researcher to account for unobserved heterogeneity and heterogeneous covariates effects, while the availability of panel data potentially allows the researcher to include fixed effects to control for some unobserved covariates. There has been little but growing work at the intersection of these two methodologies (e.g. Koenker, 2004, Geraci and Bottai, 2007, Abrevaya and Dahl, 2008, Galvao, 2008, Rosen, 2009, and Lamarche, 2010). This initial lack of attention is possibly due to a fundamental issue associated with conditional quantiles. This is, as it is the case with non-linear panel data models, standard demeaning (or differencing) techniques do not result in feasible approaches. These techniques rely on the fact that expectations are linear operators, which is not the case for conditional quantiles. This paper provides sufficient conditions under which the parameter of interest is identified for fixed T and shows that there is a simple transformation of the data that eliminates the fixed effects as $T \to \infty$, when the fixed effects are viewed as location shift variables (i.e. variables that affect all quantiles in the same way). The resulting two-step estimator is consistent and asymptotically normal when both n and T go to infinity. Also, the new estimator is extremely simple to compute and can be implemented in standard econometrics packages.

The paper is organized as follows. Section 2 presents the model. Section 3 provides an identification result based on deconvolution arguments. Section 4 introduces a two-step estimator for panel data quantile regression models. Asymptotic properties of the new estimator are presented in the same section. Section 5 includes a small Monte Carlo experiment to study the finite sample properties of the two-step estimator. Finally, Section 6 concludes. Appendix A provides proofs of results. An estimator of the covariance kernel and the bootstrap method are given in Appendix B.

2. THE MODEL

Consider the following model

$$Y_{it} = X_{it}'\theta(U_{it}) + \alpha_i, \qquad t = 1, \ldots, T, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $(Y_{it}, X_{it}) \in \mathbb{R} \times \mathbb{R}^k$ are observable variables and $(U_{it}, \alpha_i) \in \mathbb{R} \times \mathbb{R}$ are unobservable. Throughout the paper the vector X_{it} is assumed to include a constant term, i.e. $X_{it}' = (1, X_{it}^{s\prime})$ with $X_{it}^s \in \mathbb{R}^{k-1}$. The function $\tau \mapsto X'\theta(\tau)$ is assumed to be strictly increasing in $\tau \in (0, 1)$ and the parameter of interest is assumed to be $\theta(\tau)$. If α_i were observable it would follow that

$$P\big[Y_{it} \le X_{it}'\theta(\tau) + \alpha_i \mid X_i, \alpha_i\big] = \tau, \tag{2.2}$$

under the assumption that $U_{it} \sim U[0, 1]$ conditional on $X_i = (X_{i1}', \ldots, X_{iT}')'$ and α_i. This type of representation has been extensively used in the literature (e.g. Chernozhukov and Hansen, 2006, 2008). The difference with the model in equation (2.1) and the standard quantile regression model introduced by Koenker and Bassett (1978) lies in the presence of the unobserved α_i. This random variable could be arbitrarily related to the rest of the random variables in equation (2.1) (i.e. $\alpha_i = \alpha_i(U_{it}, X_i, \eta_i)$ for some i.i.d. sequence η_i) rendering condition (2.2) as not particularly useful in terms of identification. The question is under what additional conditions on (U_{it}, α_i) the parameter $\theta(\tau)$ can be identified and consistently estimated from the data.

Rosen (2009) recently showed that conditional on covariates quantile restriction alone does not identify $\theta(\tau)$. That is, let $Q_Z(\tau \mid A)$ denote the τ-quantile of a random variable Z conditional on another random variable A, let $e_{it}(\tau) \equiv X_{it}'[\theta(U_{it}) - \theta(\tau)]$, and write the model in equation (2.1) as

$$Y_{it} = X_{it}'\theta(\tau) + \alpha_i + e_{it}(\tau), \qquad Q_{e_{it}(\tau)}(\tau \mid X_i) = 0. \tag{2.3}$$

Then, the conditional quantile restriction $Q_{e_{it}(\tau)}(\tau \mid X_i) = 0$ does not have sufficient identification power.[1] Rosen (2009) then provides different assumptions, i.e. support conditions and some form of conditional independence of $e_{it}(\tau)$ across time, that (point and partially) identify $\theta(\tau)$.

Abrevaya and Dahl (2008) use the correlated random-effects model of Chamberlain (1982, 1984) as a way to get an estimator of $\theta(\tau)$. This model views the unobservable α_i as a linear projection onto the observables plus a disturbance, i.e.

$$\alpha_i\,(\tau, X_i, \eta_i) = X_i'\Lambda_T(\tau) + \eta_i.$$

The authors view the model as an approximation to the true conditional quantile and proceed to get estimates of $\theta(\tau)$ and $\Lambda_T(\tau)$ by running a quantile regression of Y_{it} on X_{it} and X_i. In cases where there is no disturbance η_i, such a regression identifies $\theta(\tau)$. However, it is immediate to see that a quantile restriction alone does not identify $\theta(\tau)$ whenever η_i is present non-trivially since the conditional behaviour of $X_{it}'\theta(U_{it}) + X_i'\Lambda_T(\tau) + \eta_i$ depends on the joint distribution of the unobservables U_{it} and η_i. This is problematic since not even a correctly specified function for $\alpha_i(\tau, X_i, \eta_i)$ helps in identifying $\theta(\tau)$, meaning that the correlated random-effects model might work poorly in many contexts. The simulations of Section 5 illustrate this point.

[1] Note that the distribution of $e_{it}(\tau)$ need not be identical across t even when U_{it} is i.i.d., but that $e_{it}(\tau)$ has the same τ-quantile for all t.

Koenker (2004) takes a different approach and treats $\{\alpha_i\}_{i=1}^n$ as parameters to be jointly estimated with $\theta(\tau)$ for q different quantiles. He proposes the penalized estimator

$$\left(\tilde{\theta}, \{\tilde{\alpha}_i\}_{i=1}^n\right) \equiv \underset{(\theta, \{\alpha_i\}_{i=1}^n)}{\operatorname{argmin}} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_{\tau_k}\left[Y_{it} - X_{it}'\theta(\tau_k) - \alpha_i\right] + \lambda \sum_{i=1}^{n} |\alpha_i|, \tag{2.4}$$

where $\rho_\tau(u) = u[\tau - I(u < 0)]$, $I(\cdot)$ denotes the indicator function, and $\lambda \geq 0$ is a penalization parameter that shrinks the $\tilde{\alpha}$s towards a common value. Solving equation (2.4) can be computationally demanding when n is large (even for $\lambda = 0$) and has the additional complication involved in the choice of λ.[2]

Finally, there is a related literature on non-separable panel data models. These type of models are flexible enough to provide quantile treatment effects (see, e.g. Chernozhukov et al., 2010, and Graham and Powell, 2010). For example, Chernozhukov et al. (2010) show that the quantile treatment effect of interest is partially identified (for fixed T) and provide bounds for those effects in the model

$$Y_{it} = g_0(X_{it}, \alpha_i, U_{it}), \qquad U_{it} \mid X_i, \alpha_i =^d U_{it'} \mid X_i, \alpha_i, \tag{2.5}$$

where X_{it} is assumed discrete. They also derive rates of shrinkage of the identified set to a point as T goes to infinity. The model in Chernozhukov et al. (2010) is more general than the one in equation (2.1) as it is non-separable in α_i and it involves weaker assumptions on the unobservable U_{it}. However, it leads to less powerful identification results and more complicated estimators.

In this context this paper contributes to the literature in two ways. The next section shows that when the model in equation (2.1) is viewed as a deconvolution model, a result from Neumann (2007) can be applied to show that $\theta(\tau)$ is identified when there are at least two time periods available and α_i has a pure location shift effect.[3] This identification result could be potentially used to construct estimators of $\theta(\tau)$ based on non-parametric estimators of conditional distribution functions. Such non-parametric estimators would rarely satisfy the end-goal of this paper, that is, to provide an easy-to-use estimator that can be implemented in standard econometric packages, would suffer from the common curse of dimensionality, and would typically involve a delicate choice of tuning parameters for their implementation.[4] Thus, when moving from identification to estimation, this paper takes a different approach and shows that there exists a simple transformation of the data that eliminates the fixed effects α_i as $T \to \infty$. The transformation leads to an extremely simple asymptotically normal estimator for $\theta(\tau)$ that can be easily computed even for very large values of n. Standard errors for this new estimator can be computed from the asymptotically normal representation.

3. IDENTIFICATION

In this section, I prove that the parameter of interest $\theta(\tau)$ is identified for $T \geq 2$ under independence restrictions and existence of moments. The intuition behind the result is quite

[2] Lamarche (2010) proposes a method to chose λ under the additional assumption that α_i and X_i are independent. Galvao (2008) further extends this idea to dynamic panels.

[3] This means that if α_i captures unobserved covariates $Z_i'\beta(\tau)$ that enter the model and are constant over time, such variables must have coefficients that are constant across τ, that is, $\beta = \beta(\tau)$ for all τ.

[4] The advantage of such non-parametric estimators would be their consistency for asymptotics in which $n \to \infty$ and T remains fixed.

simple. Letting $S_t \equiv X_t'\theta(U_t)$ (the dependence on i is omitted for convenience here), it follows from equation (2.1) that $Y_t = S_t + \alpha$ is a convolution of S_t and α conditional on X, provided α and U_t are independent conditional on X. It then follows that the conditional distributions of S_t and α can be identified from the conditional distribution of Y_t by using a deconvolution argument similar to that in Neumann (2007). This in turn results in identification of $\theta(\tau)$ after exploiting the fact that U_t is conditionally $U[0, 1]$ together with some regularity conditions.

For ease of exposition let $T = 2$ and consider the following assumption where the lower case $x = (x_1, x_2)$ denotes a realization of the random variable $X = (X_1, X_2)$.

ASSUMPTION 3.1. *Denote by $\phi_{S_t|x}$ and $\phi_{\alpha|x}$ the conditional on $X = x$ characteristic functions of the distributions $P_{S_t|x}$ and $P_{\alpha|x}$, respectively. Then (a) conditional on $X = x$ the random variables S_1, S_2 and α are independent for all $x \in \bar{\mathcal{X}}$, where $\bar{\mathcal{X}}$ denotes the support of $X = (X_1, X_2)$; (b) the set $\Gamma \equiv \{\omega : \phi_{S_t|x}(\omega 2^{-k}) \neq 0 \text{ for } t \in \{1, 2\} \text{ and } \phi_{\alpha|x}(\omega 2^{-k}) \neq 0, k = 0, 1, \ldots\}$ is dense in $\mathbb{R}$ for all $x \in \bar{\mathcal{X}}$.*

Assumption 3.1 follows Neumann (2007), but it does not impose $P_{S_1\,|\,x} = P_{S_2\,|\,x}$ for all $x \in \bar{\mathcal{X}}$, which does not hold in equation (2.1). Assumption 3.1(a) implies that α_i does not change across quantiles as α_i is independent of (U_1, U_2). Assumption 3.1(b) excludes characteristic functions that vanish on non-empty open subsets of $\mathbb{R}$ but allows the characteristic function to have countably many zeros. This includes cases ruled out in the deconvolution analysis of Li and Vuong (1998) and Evdokimov (2010), among others. More generally, any deconvolution analysis based on Kotlarski's Lemma (which has been widely applied to identify and estimate a variety of models in economics) will fail to include such cases.[5] For example, if $\alpha \sim U[-1, 1]$, or if α is discrete uniform, ϕ_α has countably many zeros and so Kotlarski's Lemma would not apply while Assumption 3.1(b) would be satisfied. Having this extra dimension of generality is important here as the distribution of α_i is left unspecified in this paper.

Assumption 3.1 implies that Y_t is a convolution of S_t and α conditional on $X = x$ and so

$$\phi_{(Y_1,Y_2)\,|\,x}(\omega_1, \omega_2) = E[\exp(i\omega' Y)] = E[\exp(i\omega_1 S_1 + i\omega_2 S_2 + i(\omega_1 + \omega_2)\alpha)\,|\,X = x] \tag{3.1}$$

$$= \phi_{S_1\,|\,x}(\omega_1)\phi_{S_2\,|\,x}(\omega_2)\phi_{\alpha|x}(\omega_1 + \omega_2). \tag{3.2}$$

LEMMA 3.1. *Suppose that $P_{S_t|x}$ and $P_{\alpha|x}$ are distributions with characteristic functions $\phi_{S_t|x}$ and $\phi_{\alpha|x}$ satisfying Assumption 3.1. Let $\tilde{P}_{\alpha|x}$ and $\tilde{P}_{S_t|x}$ be further distributions with respective characteristic functions $\tilde{\phi}_{\alpha|x}$ and $\tilde{\phi}_{S_t|x}$. If now*

$$\phi_{S_1\,|\,x}(\omega_1)\phi_{S_2\,|\,x}(\omega_2)\phi_{\alpha|x}(\omega_1 + \omega_2) = \tilde{\phi}_{S_1\,|\,x}(\omega_1)\tilde{\phi}_{S_2\,|\,x}(\omega_2)\tilde{\phi}_{\alpha|x}(\omega_1 + \omega_2), \quad \forall \omega_1, \omega_2 \in \mathbb{R}, \tag{3.3}$$

and all $x \in \bar{\mathcal{X}}$, then there exist constants $c_x \in \mathbb{R}$ such that

$$\tilde{P}_{S_t|x} = P_{(S_t - c_x)\,|\,x} \quad \text{and} \quad \tilde{P}_{\alpha|x} = P_{(\alpha + c_x)\,|\,x}. \tag{3.4}$$

[5] An extension of the result by Kotlarski for the case of characteristic functions with zeros was recently proposed by Evdokimov and White (2010) by using conditions on the derivatives of the characteristic functions.

That is, $\tilde{P}_{S_t|x}$ and $P_{S_t|x}$ for $t \in \{1, 2\}$ as well as $\tilde{P}_{\alpha|x}$ and $P_{\alpha|x}$ are equal up to a location shift.

REMARK 3.1. Lemma 3.1 extends immediately to models that are slightly more general than the one in equation (2.1). For example, consider the case where $Y_{it} = q(U_{it}, X_{it}) + \alpha_i$ and $q(\tau, x)$ is strictly increasing in τ for all $x \in \bar{\mathcal{X}}$. This model reduces to equation (2.1) if $q(U_{it}, X_{it}) = X'_{it}\theta(U_{it})$. Lemma 3.1 holds in this case by letting $S_t = q(U_t, X_t)$. Another example would be a random coefficients model where $Y_{it} = X'_{it}\theta_i(U_{it})$ and $\theta_i(U_{it}) = \theta(U_{it}) + \tilde{\alpha}_i$. This model can be written as $Y_{it} = X'_{it}\theta(U_{it}) + \alpha_{it}$ and so the additional generality relative to equation (2.1) comes from $\alpha_{it} \equiv X'_{it}\tilde{\alpha}_i$ varying across *i* and *t*. Lemma 3.1 also applies to this model conditional on the event $X_1 = X_2$.

REMARK 3.2. It is worth noting that the identification result in Lemma 3.1 also applies to the correlated random effects model in Abrevaya and Dahl (2008) where

$$Y_t = \theta(U_t)X_t + \alpha(U_t, X, \eta), \qquad \alpha(U_t, X, \eta) = \Lambda_1(U_t)X_1 + \Lambda_2(U_t)X_2 + \eta. \tag{3.5}$$

Here Assumption 3.1 must hold for $\phi_{S_t|x}$ and $\phi_{\eta|x}$ where $S_t = (\theta(U_t) + \Lambda_t(U_t))X_t + \Lambda_{-t}(U_t)X_{-t}$.

Lemma 3.1 identifies the distributions up to location. Typically, to be able to identify the entire distribution one would need to add a location assumption. In the case of quantile regression the standard cross-section specification assumes that $U \sim U[0, 1]$ independent of X. The extension of this assumption to the panel case together with some additional regularity conditions allows identification of the location of S_t conditional on $X = x$ as well as the parameter $\theta(\tau)$. This is the role of Assumption 3.2.

ASSUMPTION 3.2. *(a) $U_{it} \perp (X_i, \alpha_i)$ and $U_{it} \sim U[0, 1]$; (b) $\Omega_{UU} \equiv E[(\theta(U_{it}) - \theta_\mu)(\theta(U_{it}) - \theta_\mu)']$, where $\theta_\mu \equiv E[\theta(U_{it})]$, is non-singular with finite norm; (c) letting $X_t = (1, X_t^s)$ for $t = 1, 2$, there exists no $A \subseteq \mathbb{R}^{k-1}$ such that A has probability 1 under the distribution of $X_2^s - X_1^s$ and A is a proper linear subspace of $\mathbb{R}^{k-1}$; (d) (Y_t, X_t) have finite first moments for $t = \{1, 2\}$.*

Assumption 3.2(a) is standard in quantile regression models except that here U_{it} is also assumed independent of α_i. Assumption 3.2(b) implies that $\theta_\mu \in \mathbb{R}^k$ exists and this implies that the location of S_t is well defined. The restriction on Ω_{UU} is not used in Lemma 3.2 but it is important for the derivation of the asymptotic variance of the two-step estimator of the next section. Assumption 3.2(c) is a standard rank-type condition on the subvector of regressors that excludes the constant term. Assumption 3.2(d) is implied by Assumption 4.1 in the next section.

It is immediate to see that Assumption 3.2(a) (b) implies $E[Y_2 - Y_1 \mid X] = (X_2^s - X_1^s)'\theta_\mu^s$ and $\theta_\mu^0 = E[Y_1] - E[X_1^s]'\theta_\mu^s$, where $\theta'_\mu = (\theta_\mu^0, \theta_\mu^{s\prime})$. Assumption 3.2(c) (d) then implies that θ_μ is identified.[6] Since $E(S_t \mid X) = X'_t\theta_\mu$, the location of S_t conditional on $X = x$ is identified. Also, note that Assumption 3.2 implies

$$Q_{S_t|X=x}(t \mid x) = \theta[Q_{U_t|X=x}(t \mid x)]x = \theta(\tau)x. \tag{3.6}$$

The following Lemma then follows immediately.

[6] This follows from $A = \{x \in \mathbb{R}^{k-1} : x'(\theta_\mu^* - \theta_\mu^{**}) = 0\}$ being a proper linear subspace of $\mathbb{R}^{k-1}$ if $\theta_\mu^* \neq \theta_\mu^{**}$. In addition, if X_t does not include a constant then Assumption 3.2(c) must hold for $X_2 - X_1$.

LEMMA 3.2. *Under Assumptions 3.1 and 3.2 the location θ_μ and the function $\theta(\tau)$ for $\tau \in (0, 1)$ are identified.*

REMARK 3.3. The extension to the case $T > 2$ is straightforward under the same assumptions. The exception is Assumption 3.1(ii), which should be replaced by $\Gamma \equiv \{\omega : \phi_{S_t|x}(\omega(T-1)^{-k}) \neq 0,\ \text{for } t = 1, \ldots, T, \quad k = 0, 1, \ldots\}$ is dense in $\mathbb{R}$ for all $x \in \mathcal{X}$; see Neumann (2007).

The results of this section show that the parameter of interest $\theta(\tau)$ is *point* identified from the distribution of the observed data and Assumptions 3.1 and 3.2. The next step then is to derive a consistent estimator for this parameter and study its asymptotic properties. Section 4 presents a two-step estimator that follows a different intuition relative to the one behind the identification result, but has the virtue of being extremely simple to compute and has an asymptotically normal distribution as both n and T go to infinity.

4. TWO-STEP ESTIMATOR

The two-step estimator that I introduce in this section exploits two direct consequences of Assumption 3.2 and the fact that α_i is a location shift (Assumption 3.1(a)). The first implication is in equation (2.3), where only $\theta(\tau)$ and $e_{it}(\tau)$ depend on τ. The second implication arises by letting $u_{it} \equiv X'_{it}[\theta(U_{it}) - \theta_\mu]$ and writing a conditional mean equation for Y_{it} as follows[7]

$$Y_{it} = X'_{it}\theta_\mu + \alpha_i + u_{it}, \qquad E(u_{it} \mid X_i, \alpha_i) = 0. \tag{4.1}$$

Equation (4.1) implies that α_i is also present in the conditional mean of Y_{it}. Therefore, from equation (4.1) I can compute a $\sqrt{T}$-consistent estimator of α_i given a $\sqrt{nT}$-consistent estimator of θ_μ. This includes for example the standard within estimator of θ_μ given in equation (A.17) in Appendix A. Then, using equation (2.3) I estimate $\theta(\tau)$ by a quantile regression of the random variable $\hat{Y}_{it} \equiv Y_{it} - \hat{\alpha}_i$ on X_{it}. To be precise, the two steps are described below where I use the notation $\mathbb{E}_T(\cdot) \equiv T^{-1}\sum_{t=1}^{T}(\cdot)$ and $\mathbb{E}_{nT}(\cdot) \equiv (nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n}(\cdot)$.

Step 1. Let $\hat{\theta}_\mu$ be a $\sqrt{nT}$-consistent estimator of θ_μ. Define $\hat{\alpha}_i \equiv \mathbb{E}_T[Y_{it} - X'_{it}\hat{\theta}_\mu]$.
Step 2. Let $\hat{Y}_{it} \equiv Y_{it} - \hat{\alpha}_i$ and define the two-step estimator $\hat{\theta}(\tau)$ as:

$$\hat{\theta}(\tau) \equiv \underset{\theta\in\Theta}{\operatorname{argmin}}\, \mathbb{E}_{nT}\left[\rho_\tau\left(\hat{Y}_{it} - X'_{it}\theta\right)\right]. \tag{4.2}$$

Intuitively, the two-step estimator in equation (4.2) works because $\hat{Y}_{it} \rightsquigarrow Y^*_{it} \equiv Y_{it} - \alpha_i$ as $T \to \infty$, where $\rightsquigarrow$ denotes weak convergence. This is so because $\hat{Y}_{it} \equiv Y^*_{it} + \hat{r}_i$, where

$$\hat{r}_i \equiv (\alpha_i - \hat{\alpha}_i) = \mathbb{E}_T(X_{it})'(\hat{\theta}_\mu - \theta_\mu) - \mathbb{E}_T[Y^*_{it} - X'_{it}\theta_\mu] \to_p 0, \text{ as } T \to \infty. \tag{4.3}$$

Then, the random variable $\hat{Y}_{it}$ converges in probability, as $T \to \infty$, to the variable Y^*_{it} which implies weak convergence, $\hat{Y}_{it} \rightsquigarrow Y^*_{it}$. The next proposition shows that the two-step estimator defined in equation (4.2) is consistent and asymptotically normal under the following assumptions.

[7] Note that $E(e_{it}(\tau) \mid X_i) = X'_{it}[\theta_\mu - \theta(\tau)] \neq 0$ unless $\theta_\mu = \theta(\tau)$. Also, $\Pr(u_{it} \le 0 \mid X_i) = \Pr(X'_{it}\theta(U_{it}) \le X'_{it}\theta_\mu \mid X_i) \lesseqgtr \tau$, depending on whether $\theta(U_{it}) \lesseqgtr \theta_\mu$.

ASSUMPTION 4.1. *Let $\varphi_\tau(u) = \tau - I(u < 0)$, $I(\cdot)$ denote the indicator function, $W = (Y^*, X)$, and $g_\tau(W, \theta, r) \equiv \varphi_\tau(Y^* - X'\theta + r)X$. (a) $(Y^*_{it}, X_{it}, \alpha_i)$ are i.i.d. defined on the probability space $(\mathcal{W}, F, P)$, take values in a compact set $\mathcal{Y} \times \mathcal{X} \times \mathcal{A}$,* and $E(\alpha_i) = 0$; *(b) for all $\tau \in \mathcal{T}$, $\theta(\tau) \in int\,\Theta$, where Θ is compact and convex and $\mathcal{T}$ is a closed subinterval of* (0, 1); *(c) $Y^* \in \mathcal{Y}$ has bounded conditional on X density a.s., $\sup_{\tilde{y}\in\mathcal{Y}} f(\tilde{y}) < K$, and $\Pi(\theta, \tau, r) \equiv E[g_\tau(W, \theta, r)]$ has Jacobian matrix $J_1(\theta, \tau, r) = \frac{\partial}{\partial\theta'}\Pi(\theta, \tau, r)$ that is continuous and has full rank uniformly over $\Theta \times \mathcal{T} \times \mathcal{R}$, and $J_2(\theta, \tau, r) = \frac{\partial}{\partial r}\Pi(\theta, \tau, r)$ is uniformly continuous over $\Theta \times \mathcal{T} \times \mathcal{R}$.*

Assumption 4.1 imposes similar conditions to those in Chernozhukov and Hansen (2006). Note that 4.1(a) imposes i.i.d. on the unobserved variable Y^*_{it} and not on Y_{it}. Condition 4.1(c) is used for asymptotic normality. Finally, in order to derive the expression for the covariance kernel of the limiting Gaussian process provided in equation (4.7), I use the following assumption on the preliminary estimator $\hat{\theta}_\mu$.

ASSUMPTION 4.2. *The preliminary estimator $\hat{\theta}_\mu$ admits the expansion*

$$\sqrt{nT}(\hat{\theta}_\mu - \theta_\mu) = \sqrt{nT}\mathbb{E}_{nT}(\psi_{it}) + o_p(1), \tag{4.4}$$

where ψ_{it} is an i.i.d. sequence of random variables with $E[\psi_{it}] = 0$ and finite $\Omega_{\psi\psi} = E[\psi_{it}\psi'_{it}]$.

THEOREM 4.1. *Let $n/T^s \to 0$ for some $s \in (1, \infty)$. Under Assumptions 3.2, 4.1 and 4.2*

$$\sup_{\tau\in\mathcal{T}} \|\hat{\theta}(\tau) - \theta(\tau)\| \to_p 0,$$

and

$$\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot)) = [-J_1(\cdot)]^{-1}\sqrt{nT}\mathbb{E}_{nT}\{\varphi_\tau(\varepsilon_{it}(\tau))X_{it} + J_2(\cdot)\xi_{it}\} + o_p(1), \tag{4.5}$$

$$\rightsquigarrow \mathbb{G}(\cdot) \qquad in\ \ell^\infty(\mathcal{T}), \tag{4.6}$$

where $\varepsilon_{it}(\tau) \equiv Y^*_{it} - X'_{it}\theta(\tau)$, $\xi_{it} \equiv \mu'_X\psi_{it} - u_{it}$, $u_{it} \equiv Y^*_{it} - X'_{it}\theta_\mu$, $\mu_X \equiv E[X_{it}]$, $J_1(\tau) \equiv J_1(\theta(\tau), \tau, 0)$, $J_2(\tau) \equiv J_2(\theta(\tau), \tau, 0)$, $\mathbb{G}(\cdot)$ is a mean-zero Gaussian process with covariance function $E\mathbb{G}(\tau)\mathbb{G}(\tau')' = J_1(\tau)^{-1}\Psi(\tau, \tau')[J_1(\tau')^{-1}]'$, $\Psi(\tau, \tau')$ is defined in equation (4.7) below, and $\ell^\infty(\mathcal{T})$ in the set of all uniformly bounded functions on $\mathcal{T}$. The matrix $\Psi(\tau, \tau')$ is given by

$$\Psi(\tau, \tau') = S(\tau, \tau') + J_2(\tau)\Omega_{\xi g}(\tau') + \Omega_{g\xi}(\tau)J_2(\tau') + J_2(\tau)\Omega_{\xi\xi}J_2(\tau')', \tag{4.7}$$

where $S(\tau, \tau') \equiv (\min\{\tau, \tau'\} - \tau\tau')E(XX')$, $\Omega_{g\xi}(\tau) \equiv E[g_\tau(W, \theta(\tau))\xi]$, and $\Omega_{\xi\xi} \equiv E[\xi^2]$.

The asymptotic expansion for $\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot))$ presented in Theorem 4.1 has two terms. The first term is the standard term in quantile regressions while the second term captures the fact α_i is being estimated by $\hat{\alpha}_i$. Assumption 4.2 is not necessary for convergence to a limit Gaussian process but it is used to derive the expression of $\Psi(\tau, \tau')$ in equation (4.7). This is a mild assumption as, for example, the usual within estimator of θ_μ satisfies Assumption 4.2 under Assumptions 3.2 and 4.1 (see Lemma A.4 in Appendix A).

Appendix B contains expressions for estimating each element entering the covariance function $J_1(\tau)^{-1}\Psi(\tau, \tau')[J_1(\tau')^{-1}]'$ that can be used to make inference on $\hat{\theta}(\tau)$. Alternatively, I conjecture that the standard i.i.d. bootstrap on (Y_i, X_i) would work in this context and therefore provide a different approach to inference (Appendix B). The proof of this conjecture is however beyond the scope of this paper and the supporting evidence is limited to the simulations of the Section 5.

5. SIMULATIONS

To illustrate the performance of the two-step estimator I conduct a small simulation study. Tables 1 and 2 summarize the results by reporting percentage bias (%Bias) and mean squared error (MSE) for each estimator considered. The simulated model is

$$\begin{aligned} Y_{it} &= (\epsilon_{it} - 1) + \epsilon_{it} X_{it} + \alpha_i, \\ \alpha_i &= \gamma(X_{i1} + \cdots + X_{iT} + \eta_i) - E(\alpha_i), \end{aligned} \tag{5.1}$$

Table 1. Bias and MSE for Model 5.1: $\gamma = 2$ and 0.25.

		$n = 100$				$n = 5000$			
		$\theta(\cdot) \sim N(2, 1)$, $\theta_{[1]}(\tau) = 1.3255$				$\theta(\cdot) \sim N(2, 1)$, $\theta_{[1]}(\tau) = 1.3255$			
	Estimator	QR	CRE	INFE	2-STEP	QR	CRE	INFE	2-STEP
$T = 5$	%Bias	1.7569	0.2157	0.0068	0.1494	1.7444	0.2074	0.0004	0.1496
	MSE	5.8220	0.2794	0.1032	0.1473	5.3543	0.0798	0.0020	0.0414
$T = 10$	%Bias	1.7660	0.2127	0.0025	0.0793	1.7687	0.2063	−0.0020	0.0757
	MSE	5.7557	0.1639	0.0494	0.0605	5.5021	0.0765	0.0010	0.0111
$T = 20$	%Bias	1.7750	0.1903	0.0032	0.0377	1.8084	0.2070	−0.0001	0.0397
	MSE	5.7678	0.1044	0.0228	0.0264	5.7506	0.0762	0.0005	0.0032
		$\theta(\cdot) \sim \exp(1) + 2$, $\theta_{[1]}(\tau) = 2.2877$				$\theta(\cdot) \sim \exp(1) + 2$, $\theta_{[1]}(\tau) = 2.2877$			
	Estimator	QR	CRE	INFE	2-STEP	QR	CRE	INFE	2-STEP
$T = 5$	%Bias	1.0202	0.1291	−0.0003	0.0976	1.0285	0.1367	−0.0000	0.0961
	MSE	5.8106	0.2066	0.0174	0.1003	5.5434	0.1002	0.0003	0.0493
$T = 10$	%Bias	1.0310	0.1184	−0.0002	0.0487	1.0393	0.1368	−0.0001	0.0489
	MSE	5.8306	0.1198	0.0087	0.0286	5.6593	0.0990	0.0002	0.0128
$T = 20$	%Bias	1.0524	0.0919	0.0008	0.0215	1.0593	0.1361	−0.0001	0.0209
	MSE	6.0432	0.0641	0.0045	0.0087	5.8777	0.0975	0.0001	0.0024
		$\theta(\cdot) \sim Mixture$, $\theta_{[1]}(\tau) = 1.3097$				$\theta(\cdot) \sim Mixture$, $\theta_{[1]}(\tau) = 1.3097$			
	Estimator	QR	CRE	INFE	2-STEP	QR	CRE	INFE	2-STEP
$T = 5$	%Bias	2.0565	0.5369	0.0190	0.2561	2.1061	0.5370	0.0012	0.1897
	MSE	7.6837	0.7855	0.5115	0.4963	7.5625	0.4965	0.0036	0.0682
$T = 10$	%Bias	2.1221	0.5683	0.0026	0.0424	2.1433	0.5413	0.0023	0.0280
	MSE	7.9763	0.6679	0.1450	0.2088	7.8295	0.5016	0.0018	0.0046
$T = 20$	%Bias	2.1770	0.6010	−0.0039	−0.0008	2.1750	0.5412	−0.0013	−0.0032
	MSE	8.3200	0.6744	0.0536	0.0763	8.0616	0.5001	0.0009	0.0013

Note: 1000 MC replications. QR: standard quantile regression estimator; CRE: correlated random effect estimator; INFE: infeasible estimator, 2-STEP: two-step estimator from 4.2. All regressions include an intercept term.

Table 2. Bias and MSE for Model 5.1: $\gamma = 2$ and 0.90.

		$n = 100$				$n = 5000$			
		$\theta(\cdot) \sim N(2, 1), \theta_{[1]}(\tau) = 3.2815$				$\theta(\cdot) \sim N(2, 1), \theta_{[1]}(\tau) = 3.2815$			
	Estimator	QR	CRE	INFE	2-STEP	QR	CRE	INFE	2-STEP
$T = 5$	%Bias	0.4069	−0.1697	−0.0065	−0.1223	0.4321	−0.1575	−0.0002	−0.1144
	MSE	2.4048	0.6455	0.1509	0.3162	2.0220	0.2743	0.0032	0.1439
$T = 10$	%Bias	0.4039	−0.1547	−0.0069	−0.0645	0.4079	−0.1594	−0.0006	−0.0603
	MSE	2.1894	0.3995	0.0763	0.1228	1.8001	0.2768	0.0015	0.0406
$T = 20$	%Bias	0.3777	−0.1307	0.0013	−0.0280	0.3833	−0.1579	0.0001	−0.0301
	MSE	1.8410	0.2466	0.0393	0.0479	1.5884	0.2700	0.0008	0.0105
		$\theta(\cdot) \sim \exp(1) + 2, \theta_{[1]}(\tau) = 4.3026$				$\theta(\cdot) \sim \exp(1) + 2, \theta_{[1]}(\tau) = 4.3026$			
	Estimator	QR	CRE	INFE	2-STEP	QR	CRE	INFE	2-STEP
$T = 5$	%Bias	0.3120	−0.1517	−0.0015	−0.0814	0.3134	−0.1366	−0.0002	−0.0804
	MSE	2.5781	0.9598	0.4730	0.4583	1.8333	0.3566	0.0095	0.1264
$T = 10$	%Bias	0.2911	−0.1442	0.0012	−0.0407	0.2984	−0.1361	−0.0005	−0.0417
	MSE	2.0493	0.6100	0.2337	0.2444	1.6585	0.3479	0.0049	0.0363
$T = 20$	%Bias	0.2748	−0.1261	0.0008	−0.0224	0.2823	−0.1371	−0.0003	−0.0216
	MSE	1.8136	0.4023	0.1166	0.1181	1.4834	0.3501	0.0023	0.0107
		$\theta(\cdot) \sim Mixture, \theta_{[1]}(\tau) = 3.335$				$\theta(\cdot) \sim Mixture, \theta_{[1]}(\tau) = 3.335$			
	Estimator	QR	CRE	INFE	2-STEP	QR	CRE	INFE	2-STEP
$T = 5$	%Bias	0.4921	−0.0843	−0.0040	−0.0984	0.4922	−0.0948	0.0004	−0.0958
	MSE	3.2573	0.2499	0.0196	0.1751	2.7050	0.1047	0.0004	0.1034
$T = 10$	%Bias	0.4779	−0.0808	−0.0005	−0.0584	0.4784	−0.0949	0.0009	−0.0568
	MSE	2.9631	0.1275	0.0101	0.0590	2.5542	0.1021	0.0002	0.0363
$T = 20$	%Bias	0.4556	−0.0732	−0.0002	−0.0319	0.4585	−0.0953	0.0006	−0.0325
	MSE	2.6537	0.0784	0.0048	0.0191	2.3448	0.1018	0.0001	0.0119

Note: 1000 MC replications. QR: standard quantile regression estimator; CRE: correlated random effect estimator; INFE: infeasible estimator, 2-STEP: two-step estimator from 4.2. All regressions include an intercept term.

where $X_{it} \sim Beta(1, 1)$, $\eta_i \sim N(0, 1)$ and the distribution of ϵ_{it} changes with the model specification. In Model 1, $\epsilon_{it} \sim N(2, 1)$; in Model 2, $\epsilon_{it} \sim \exp(1) + 2$; while in Model 3, $\epsilon_{it} \sim B_{it}N(1, .1) + (1 - B_{it})N(3, .1)$ with $B_{it} \sim Bernoulli(p = 0.3)$. The conditional τth quantiles are given by $\theta_{[0]}(\tau) + \theta_{[1]}(\tau)X + \alpha$. In Model 1, for example, $\theta_{[0]}(\tau) = \Phi^{-1}(\tau) + 1$ and $\theta_{[1]}(\tau) = \Phi^{-1}(\tau) + 2$ for $\Phi^{-1}(\tau)$ the inverse of the normal CDF evaluated at τ. For all models I set $n = \{100, 5000\}$, $T = \{5, 10, 20\}$, $\tau = \{0.25, 0.90\}$, $\gamma = 2$ and consider four estimators. These are a standard quantile regression estimator of Y_{it} on X_{it} (QR), a correlated random effect estimator (CRE) from Abrevaya and Dahl (2008), an infeasible estimator of $Y^*_{it} \equiv Y_{it} - \alpha_i$ on X_{it} (INFE) and the two-step estimator (2-STEP) from Section 4.[8] Note that QR and CRE are inconsistent estimators. Finally, I set the number of Monte Carlo simulations to 1000.

[8] I use the within estimator of equation (A.17) as first step estimator.

Table 3. Bias and MSE for Model 5.1: $\gamma = 2$ and $n = 100$.

		Model 1			Model 2			Model 3		
$\tau = 0.25$	Estimator	%Bias	MSE	Time	%Bias	MSE	Time	%Bias	MSE	Time
$T = 5$	2-STEP	0.1494	0.1469	0.009	0.0976	0.1003	0.009	0.2561	0.4963	0.009
	KOEN	0.1084	0.1473	0.133	−0.0014	0.0219	0.136	0.5038	0.5565	0.135
$T = 10$	2-STEP	0.0722	0.0579	0.017	0.0487	0.0286	0.018	0.0424	0.2088	0.019
	KOEN	0.0504	0.0545	0.258	−0.0061	0.0086	0.254	0.3488	0.2841	0.284
$T = 20$	2-STEP	0.0377	0.0264	0.016	0.0215	0.0087	0.020	−0.0008	0.0763	0.015
	KOEN	0.0223	0.0268	0.240	−0.0024	0.0048	0.241	0.1924	0.1129	0.233
$\tau = 0.90$	Estimator	%Bias	MSE	Time	%Bias	MSE	Time	%Bias	MSE	Time
$T = 5$	2-STEP	−0.1223	0.3162	0.009	−0.0814	0.4583	0.013	−0.0984	0.1751	0.011
	KOEN	−0.1860	0.5377	0.137	−0.2382	1.2574	0.187	−0.0617	0.0691	0.158
$T = 10$	2-STEP	−0.0645	0.1228	0.019	−0.0407	0.2444	0.018	−0.0584	0.0590	0.022
	KOEN	−0.0555	0.1371	0.262	−0.0850	0.3708	0.239	−0.0177	0.0180	0.270
$T = 20$	2-STEP	−0.0280	0.0479	0.021	−0.0224	0.1181	0.024	−0.0319	0.0191	0.026
	KOEN	−0.0280	0.0543	0.241	−0.0484	0.1594	0.253	−0.0081	0.0068	0.275

Note: 1000 MC replications. KOEN: Koenker's estimator for panel data quantile regression, 2-STEP: two-step estimator from 4.2. All regressions include an intercept term. Time is reported in seconds.

Tables 1 and 2 show equivalent patterns. CRE has a larger bias than 2-STEP and its bias does not improve as T grows.[9] 2-STEP does show a bias that decreases as T grows. This is consistent with the analysis presented in Sections 2 and 4. Also, it is worth noticing that the bias of 2-STEP is not affected as N increases and T remains fixed. Comparisons based on MSE arrive to similar conclusions. Finally, additional simulations not reported show that these results also hold for other quantiles, and that the two-step estimator performs similarly when α_i is specified as a non-linear function of X_i and η_i.

Table 3 reports bias, MSE and computational time for 2-STEP and KOEN, the estimator proposed by Koenker (2004), see equation (2.4). Only the case $n = 100$ is shown since KOEN had problems handling the big matrices for the case $n = 5000$. The results are quite mixed. 2-STEP performs better than KOEN in about half of the cases. However, the worst performance of 2-STEP (bias: 25%, MSE: 0.50) is better than that of KOEN (bias: 50%, MSE: 1.26). Note also that 2-STEP is about 15 times faster than KOEN.

Finally, Table 4 reports standard errors and 95% confidence intervals computed using the formulas provided in Appendix B. The coverage is very close to the nominal level when $T = 20$ and below the nominal level for $T = 5$. It is worth noting that the coverage is expected to deteriorate in two circumstances. Given a value of n, a smaller T implies a larger finite sample bias and so a finite sample distribution centred further away from the truth. In addition, given a value of T, a larger value of n keeps the finite sample bias unaffected but implies a finite sample distribution that is more concentrated about the wrong place. However, even for a case with 12% of bias (model 1, $T = 5$), the actual coverage levels are about 85%, which look very decent for such small values of T and bias above 10%.

[9] The performance of CRE depends on the parameter γ, so as γ grows its performance deteriorates.

Table 4. Standard Errors and 95% Confidence Intervals for Model 5.1: $\gamma = 2$ and $n = 100$.

		$\tau = 0.25$				$\tau = 0.90$		
		$T = 5$	$T = 10$	$T = 20$		$T = 5$	$T = 10$	$T = 20$
Model 1	Estimator	1.5329	1.4194	1.3719	Estimator	2.9057	3.0995	3.1822
	Asy SE	0.2934	0.2162	0.1555	Asy SE	0.3503	0.2620	0.1901
	Boot SE	0.2868	0.2131	0.1537	Boot SE	0.3689	0.2713	0.1944
	Asy Cov	0.8540	0.9230	0.9420	Asy Cov	0.7600	0.8920	0.9240
	Boot Cov	0.8320	0.9030	0.9340	Boot Cov	0.7930	0.8870	0.9240
Model 2	Estimator	2.5028	2.4046	2.3369	Estimator	3.9575	4.1257	4.2009
	Asy SE	0.1961	0.1231	0.0773	Asy SE	0.5336	0.4290	0.3254
	Boot SE	0.1904	0.1181	0.0738	Boot SE	0.5697	0.4447	0.3335
	Asy Cov	0.7760	0.8240	0.8910	Asy Cov	0.8620	0.9090	0.9220
	Boot Cov	0.7260	0.7980	0.8910	Boot Cov	0.8720	0.9090	0.9220
Model 3	Estimator	1.6305	1.3778	1.3110	Estimator	3.0058	3.1488	3.2298
	Asy SE	0.4751	0.3943	0.2630	Asy SE	0.2218	0.1320	0.0822
	Boot SE	0.5364	0.4494	0.2918	Boot SE	0.2333	0.1368	0.0842
	Asy Cov	0.8630	0.9290	0.9460	Asy Cov	0.6520	0.7050	0.7400
	Boot Cov	0.8720	0.9310	0.9460	Boot Cov	0.6800	0.7050	0.7400

Note: 1000 MC replications. Asy SE: asymptotic standard errors. Boot SE: bootstrap standard errors. Asy Cov: Coverage of the asymptotic confidence interval. Boot Cov: Coverage of the Bootstrap percentile interval.

6. DISCUSSION

This paper provided an identification result for quantile regression in panel data models and introduced a two-step estimator that is attractive for its computational simplicity. There are many issues that remain to be investigated. First, several panels available have a short time span and therefore approximations taking T to infinity might result in poor approximations for those cases. However, a computationally simple estimator that works for fixed T and large N is extremely challenging since, even under the assumption that α_i is independent of the rest of the variables of the model, we would still have to face similar problems to those discussed in Sections 2 and 3. Second, the assumption that α_i does not depend on the quantiles restricts the type of unobserved heterogeneity that the model can handle. Improvements in any of these directions are important for future research.

ACKNOWLEDGMENTS

This paper was previously circulated under the title ‘A Note on Quantile Regression for Panel Data Models.’ I am deeply grateful to Jack Porter and Bruce Hansen for thoughtful discussions. I would also like to thank the Editor and two anonymous referees whose comments have led to an improved version of this paper.

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*, 379–97.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics 18*, 5–46.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics, Volume 2*, 1247–318. New York: Academic Press.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2010). Average and quantile effects in nonseparable panel models. Working paper, M.I.T., Cambridge.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics 123*, 491–525.

Chernozhukov, V. and C. Hansen (2008). Instrumental variable quantile regression: a robust inference approach. *Journal of Econometrics 142*, 379–98.

Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working paper, Princeton University.

Evdokimov, K. and H. White (2010). An extension of a lemma of kotlarski. Manuscript, Princeton University, New Jersey.

Galvao, A. (2008). Quantile regression for dynamic panel data with fixed effects. Working paper, University of Illinois. Urbana-Champaign.

Geraci, M. and M. Bottai (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics 8*, 140–54.

Graham, B. S. and J. Powell (2010). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. Working paper, New York University.

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

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monograph, ESM38, Cambridge: Cambridge University Press.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

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

Li, T. and Q. Vuong (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis 65*, 139–65.

Neumann, M. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis 98*, 1955–68.

Powell, J. (1991). Estimation of monotonic regression models under quantile restrictions. In W. Barnett, J. Powell, and G. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics*, 357–84. Cambridge: Cambridge University Press.

Rosen, A. (2009). Set identification via quantile restrictions in short panels. Working paper, University College, London.

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

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

APPENDIX A: PROOF OF THE LEMMAS AND THEOREMS

Throughout the Appendix I use the following notation. For $W = (Y^*, X)$,

$$g \mapsto \mathbb{E}_{nT} g(W) \equiv \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} g(W_{it}), \quad g \mapsto \mathbb{G}_{nT} g(W) \equiv \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \sum_{i=1}^{n} [g(W_{it}) - Eg(W_{it})].$$

In addition, $\rightsquigarrow$ denotes weak convergence, CMT refers to the Continuous Mapping Theorem and LLN refers to the Law of Large Numbers. The symbols o, O, o_p and O_p denote the usual order of magnitudes for non-random and random sequences.

Proof of Lemma 3.1: The proof is a simple extension of the result Lemma 2.1 from Neumann (2007). I write it here for completeness. Since $\phi_{S_t|x}$ and $\phi_{\alpha|x}$ are characteristic functions there exists an $\omega_0 > 0$ such that $\phi_{S_t|x}(\omega) \neq 0$ for $t \in \{1, 2\}$ and $\phi_{\alpha|x}(2\omega) \neq 0$ if $|\omega| \leq \omega_0$ (In this case, ω_0 might depend on x but we omit this dependence for simplicity). For $|\omega| \leq 2\omega_0$ define,

$$g_{\alpha|x}(\omega) = \tilde{\phi}_{\alpha|x}(\omega)/\phi_{\alpha|x}(\omega),$$

a continuous complex function which equals 1 at 0. It follows from equation (3.3), for $\omega_1, \omega_2 \in [-\omega_0, \omega_0]$, that

$$\begin{aligned} \frac{\phi_{\alpha|x}(\omega_1+\omega_2)}{\phi_{\alpha|x}(\omega_1)\phi_{\alpha|x}(\omega_2)} &= \frac{\phi_{\alpha|x}(\omega_1+\omega_2)\phi_{S_1\,|\,x}(\omega_1)\phi_{S_2\,|\,x}(\omega_2)}{\phi_{\alpha|x}(\omega_1+0)\phi_{S_1\,|\,x}(\omega_1)\phi_{S_2\,|\,x}(0)\phi_{\alpha|x}(0+\omega_2)\phi_{S_1\,|\,x}(0)\phi_{S_2\,|\,x}(\omega_2)} \\ &= \frac{\tilde{\phi}_{\alpha|x}(\omega_1+\omega_2)\tilde{\phi}_{S_1\,|\,x}(\omega_1)\tilde{\phi}_{S_2\,|\,x}(\omega_2)}{\tilde{\phi}_{\alpha|x}(\omega_1+0)\tilde{\phi}_{S_1\,|\,x}(\omega_1)\tilde{\phi}_{S_2\,|\,x}(0)\tilde{\phi}_{\alpha|x}(0+\omega_2)\tilde{\phi}_{S_1\,|\,x}(0)\tilde{\phi}_{S_2\,|\,x}(\omega_2)} \\ &= \frac{\tilde{\phi}_{\alpha|x}(\omega_1+\omega_2)}{\tilde{\phi}_{\alpha|x}(\omega_1)\tilde{\phi}_{\alpha|x}(\omega_2)}, \end{aligned} \tag{A.1}$$

which implies

$$g_{\alpha|x}(\omega_1+\omega_2) = g_{\alpha|x}(\omega_1)g_{\alpha|x}(\omega_2) \quad \forall \omega_1, \omega_2 \in [-\omega_0, \omega_0].$$

The unique solution to this equation satisfying $g_{\alpha|x}(0) = 1$ and $g_{\alpha|x}(-\omega) = \overline{g_{\alpha|x}(\omega)}$ (a Hermitian function) is $g_{\alpha|x}(\omega) = e^{ic\omega}$, from some real c. Therefore we conclude that,

$$\tilde{\phi}_{\alpha|x}(\omega) = e^{ic\omega}\phi_{\alpha|x}(\omega) \quad \forall \omega \in [-2\omega_0, 2\omega_0]. \tag{A.2}$$

Furthermore, equation (3.3) yields that for $\omega_2 = 0$,

$$\phi_{S_1\,|\,x}(\omega_1)\phi_{\alpha|x}(\omega_1) = e^{ic\omega_1}\phi_{\alpha|x}(\omega_1)\tilde{\phi}_{S_1\,|\,x}(\omega_1), \quad \forall \omega_1 \in [-2\omega_0, 2\omega_0],$$

so that,

$$\tilde{\phi}_{S_1\,|\,x}(\omega) = e^{-ic\omega}\phi_{S_1\,|\,x}(\omega) \quad \forall \omega \in [-2\omega_0, 2\omega_0]. \tag{A.3}$$

Setting $\omega_1 = 0$ it follows that equation (A.3) holds for S_2 as well. Now it remains to extend these results to the whole real line. Let $\omega \in \Gamma$ be arbitrary. We obtain, analogously to equation (A.1), that

$$\frac{\phi_{\alpha|x}(\omega)}{(\phi_{\alpha|x}(\omega/2))^2} = \frac{\tilde{\phi}_{\alpha|x}(\omega)}{(\tilde{\phi}_{\alpha|x}(\omega/2))^2} \quad \text{and} \quad \frac{\phi_{\alpha|x}(\omega)}{(\phi_{\alpha|x}(\omega 2^{-k}))^{2^k}} = \frac{\tilde{\phi}_{\alpha|x}(\omega)}{(\tilde{\phi}_{\alpha|x}(\omega 2^{-k}))^{2^k}},$$

after iterating. Using this equation with a k large enough such that $|\omega 2^{-k}| \leq 2\omega_0$ we conclude from equation (A.2) that

$$\tilde{\phi}_{\alpha|x}(\omega) = \phi_{\alpha|x}(\omega)\left(\frac{\tilde{\phi}_{\alpha|x}(\omega 2^{-k})}{\phi_{\alpha|x}(\omega 2^{-k})}\right)^{2^k} = e^{ic\omega}\phi_{\alpha|x}(\omega).$$

Since Γ is dense in $\mathbb{R}$ and $\{\omega : \tilde{\phi}_{\alpha|x}(\omega) = e^{ic\omega}\phi_{\alpha|x}(\omega)\}$ is a closed set, we conclude that $\tilde{\phi}_{\alpha|x}(\omega) = e^{ic\omega}\phi_{\alpha|x}(\omega)$, $\forall\omega \in \mathbb{R}$, this is, $\tilde{P}_{\alpha|x} = P_{(\alpha+c)|x}$. This implies, again by equation (3.3) that $\tilde{\phi}_{S_t|x}(\omega) = e^{-ic\omega}\phi_{S_t|x}(\omega)$, $\forall\omega \in \mathbb{R}$, which yields $\tilde{P}_{S_t|x} = P_{(S_t-c)|x}$ for $t \in \{1, 2\}$. □

LEMMA A.1. *Under Assumptions 3.2 and 4.1, the following statements are true. (a)* $\mathbb{G}_{nT}g_{\cdot}(W, \theta(\cdot)) \rightsquigarrow \mathbb{G}_1^*(\cdot)$ *in* $\ell^\infty(\mathcal{T})$, *where* $\mathbb{G}_1^*$ *is a Gaussian process with covariance function* $E\mathbb{G}_1^*(\tau)\mathbb{G}_1^*(\tau')' = (\min\{\tau, \tau'\} - \tau\tau')E(XX')$. *(b) If* $\sup_{\tau\in\mathcal{T}} \|\widetilde{\theta}(\tau) - \theta(\tau)\| = o_p(1)$ and $\max_{i\leq n} |\widetilde{r}_i| = o_p(1)$, *then*

$$\sup_{\tau\in\mathcal{T}} \|\mathbb{G}_{nT}g_\tau(W, \widetilde{\theta}(\tau), \widetilde{r}_i) - \mathbb{G}_{nT}g_\tau(W, \theta(\tau), 0)\| = o_p(1).$$

Proof: The proof follows by similar arguments to those in Lemma B.2 of Chernozhukov and Hansen (2006) after noticing that the class of functions, $\mathcal{H} = \{h = (\theta, \tau, r) \mapsto \varphi_\tau(Y^* - X'\theta + r)X, \ \ \theta \in \Theta, \ \tau \in \mathcal{T}, r \in \mathcal{R}\}$ is Donsker (it is formed by taking products and sums of bounded Donsker classes) by Theorem 2.10.6 in van der Vaart and Wellner (1996). □

LEMMA A.2. *Under Assumptions 3.2 and 4.1,* $\max_{i\leq n} |\hat{r}_i| \equiv \max_{i\leq n} |\alpha_i - \hat{\alpha}_i| \to 0$ *provided* $\frac{n}{T^s} \to 0$ *for some* $s \in (1, \infty)$.

Proof: Let $u_{it} \equiv Y_{it}^* - X_{it}'\theta_\mu$ and note that by the triangle inequality,

$$|\alpha_i - \hat{\alpha}_i| \leq |\mathbb{E}_T(X_{it})|'|\hat{\theta}_\mu - \theta_\mu| + |\mathbb{E}_T(u_{it})|.$$

X_{it} and Y_{it} have compact support so that $\max_{i\leq n}|X_{it}| \leq C_x < \infty$ and $\max_{i\leq n}|u_{it}| \leq C_z < \infty$. It is immediate then that

$$\max_{i\leq n} |\mathbb{E}_T(X_{it})|'|\hat{\theta}_\mu - \theta_\mu| = o_p(1).$$

Since $E(u_{it}) = 0$ and $E(|u_{it}|^{2s}) \leq C_z^{2s} < \infty$ for all $s \in (1, \infty)$, it follows from the Markov inequality that for any $\eta > 0$, $\Pr(|\mathbb{E}_T(u_{it})| > \eta) = O(T^{-s})$ and then

$$\Pr\left(\max_{i\leq n} |\mathbb{E}_T(u_{it})| > \eta\right) \leq n\Pr(|\mathbb{E}_T(u_{it})| > \eta) = O(n/T^s) = o(1).$$

□

Proof of Theorem 4.1: Consistency. Define the following two criterion functions,

$$Q_{nT}(\theta, \tau) = \mathbb{E}_{nT}\left[\rho_\tau\left(\hat{Y}_{it} - X_{it}'\theta\right)\right] \quad \text{and} \quad Q(\theta, \tau) = E\left[\rho_\tau\left(Y_{it}^* - X_{it}'\theta\right)\right].$$

The first step shows that $Q_{nT}(\theta, \tau)$ converge uniformly to $Q(\theta, \tau)$. To this end, note that since $\hat{Y}_{it} \rightsquigarrow Y_{it}^*$, it follows from van der Vaart (1998, Lemma 2.2) that

$$\left|E\left[\rho_\tau\left(\hat{Y}_{it} - X_{it}'\theta\right)\right] - E\left[\rho_\tau\left(Y_{it}^* - X_{it}'\theta\right)\right]\right| \to 0, \text{ as } T \to \infty,$$

since $\rho_\tau(\cdot)$ is a bounded Lipschitz function (it is bounded due to 4.1(a) and 4.1(b)). Due to the compactness of $\Theta \times \mathcal{T}$ and the continuity of $E[\rho_\tau(\hat{Y}_{it} - X_{it}'\theta)]$ and $E[\rho_\tau(Y_{it}^* - X_{it}'\theta)]$ implied by 4.1(c), the above convergence is also uniform,

$$\sup_{(\theta,\tau)\in\Theta\times\mathcal{T}} |E[\rho_\tau(\hat{Y} - X'\theta)] - E[\rho_\tau(Y^* - X'\theta)]| \to 0, \text{ as } T \to \infty. \tag{A.4}$$

Next note that functions in the class $\mathcal{F} = [(\theta, \tau) \longmapsto \rho_\tau(Y - X'\theta)]$ are bounded, uniformly Lipschitz over $\Theta \times \mathcal{T}$ and form a Donsker class. This also means that $\mathcal{F}$ is Glivenko-Cantelli so that,

$$\sup_{(\theta,\tau)\in\Theta\times\mathcal{T}} \sup|\mathbb{E}_{nT}[\rho_\tau(\hat{Y} - X'\theta)] - E[\rho_\tau(\hat{Y} - X'\theta)]| \to_p 0, \text{ as } n, T \to \infty. \tag{A.5}$$

It follows from equations (A.4) and (A.5) that $Q_{nT}(\theta, \tau)$ converges uniformly to $Q(\theta, \tau)$ as both n and T go to infinity. Under Assumption 4.1(c) $\theta(\tau)$ uniquely solves $\theta(\tau) \equiv \arg\inf_{\theta\in\Theta} E[\rho_\tau(Y^* - X'\theta)]$ and $Q(\theta, \tau)$ is continuous over $\Theta \times \mathcal{T}$ so that

$$\sup_{\tau\in\mathcal{T}} \|\hat{\theta}(\tau) - \theta(\tau)\| \to_p 0,$$

after invoking Chernozhukov and Hansen (2006, Lemma B.1).

Asymptotic Normality. From the properties of standard quantile regression it follows that $\sqrt{nT}\mathbb{E}_{nT} g_\cdot(W, \hat{\theta}(\cdot), \hat{r}_i)$ is $o_p(1)$, and then the following expansion is valid,

$$\begin{aligned} o_p(1) &= \mathbb{G}_{nT} g_\cdot(W, \hat{\theta}(\cdot), \hat{r}_i) + \frac{1}{\sqrt{nT}} \sum_{t=1}^{T}\sum_{i=1}^{n} Eg_\cdot(W, \hat{\theta}(\cdot), \hat{r}_i) \\ &\overset{(1)}{=} \mathbb{G}_{nT} g_\cdot(W, \theta(\cdot)) + o_p(1) + \frac{1}{\sqrt{nT}} \sum_{t=1}^{T}\sum_{i=1}^{n} Eg_\cdot(W, \hat{\theta}(\cdot), \hat{r}_i) \qquad in\ \ell^\infty(\mathcal{T}). \end{aligned}$$

Here $=^{(1)}$ follows from Lemma A.1. Now, expand $Eg_\cdot(W, \hat{\theta}(\cdot), \hat{r}_i)$.

$$\begin{aligned} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T}\sum_{i=1}^{n} Eg_\cdot(W, \hat{\theta}(\cdot), \hat{r}_i) &= \frac{1}{\sqrt{nT}} \sum_{t=1}^{T}\sum_{i=1}^{n} \left\{J_1\left(\theta^*(\cdot), \cdot, r_i^*\right)(\hat{\theta}(\cdot) - \theta(\cdot)) + J_2\left(\theta^*(\cdot), \cdot, r_i^*\right)\hat{r}_i\right\} \\ &= J_1(\cdot)\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot)) + J_2(\cdot)\frac{1}{\sqrt{nT}} \sum_{t=1}^{T}\sum_{i=1}^{n} \hat{r}_i + o_p(1), \end{aligned}$$

where θ^* is on the line connecting $\hat{\theta}(\tau)$ and $\theta(\tau)$ for each τ and r_i^* is on the line connecting 0 and $\hat{r}_i$. The second equality follows from

$$\sup_{i\le n,\tau\in\mathcal{T}} \left|J_k\left(\theta^*(\cdot), \tau, r_i^*\right) - J_k(\theta(\tau), \tau, 0)\right| = o_p(1), \quad \text{for } k = 1, 2, \tag{A.6}$$

which in turn follows from the uniform continuity assumption, and the fact that $\max_{i\le n} |\hat{r}_i| = o_p(1)$ by Lemma A.2. Solving for $\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot))$,

$$\begin{aligned} \sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot)) &= [-J_1(\cdot)]^{-1}\, \mathbb{G}_{nT} g_\tau(W, \theta(\cdot)) + [-J_1(\cdot)]^{-1}\, J_2(\cdot)\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{T}\hat{r}_i + o_p(1) \\ &\rightsquigarrow \mathbb{G}(\cdot), \quad \text{in } \ell^\infty(\mathcal{T}), \end{aligned}$$

where $\mathbb{G}(\cdot)$ is a gaussian process with covariance kernel $J_1(\tau)^{-1}\Psi(\tau, \tau')[J_1(\tau')^{-1}]'$, where $\Psi(\tau, \tau')$ is defined in equation (A.10). This follows from the first term converging to a Gaussian process by Lemma A.1, and the second term

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{T}\hat{r}_i &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{T}\left[\mathbb{E}_T(X_{it})'(\hat{\theta}_\mu - \theta_\mu) - \mathbb{E}_T\left[Y_{it}^* - X_{it}'\theta_\mu\right]\right], \\ &= \mathbb{E}_{nT}(X_{it})'\sqrt{nT}(\hat{\theta}_\mu - \theta_\mu) - \sqrt{nT}\mathbb{E}_{nT}\left(Y_{it}^* - X_{it}'\theta_\mu\right), \end{aligned}$$

being asymptotically normal due to $\sqrt{nT}(\hat{\theta}_\mu - \theta_\mu)$, and $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}(Y_{it}^* - X_{it}'\theta_\mu)$ both being asymptotically normal.

Covariance Kernel. We first need to derive the expression for $\Psi(\tau, \tau')$. Under Assumption 4.2 we can write the expansion $\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot))$ as

$$\begin{aligned}\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot)) &= [-J_1(\cdot)]^{-1}\left\{\mathbb{G}_{nT} g_\tau(W, \theta(\cdot)) + J_2(\cdot)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{T}\hat{r}_i\right\} + o_p(1), \\ &= [-J_1(\cdot)]^{-1}\sqrt{nT}\mathbb{E}_{nT}\{g_\cdot(W_{it}, \theta(\cdot)) + J_2(\cdot)\xi_{it}\} + o_p(1),\end{aligned} \tag{A.7}$$

where $\xi_{it} \equiv \mu_X'\psi_{it} - u_{it}$, $\mu_X \equiv E(X)$, and $u_{it} \equiv Y_{it}^* - X_{it}'\theta_\mu$. In addition,

$$J_1(\tau) = E[f_{\varepsilon(\tau)}(0 \mid X)XX'],$$

where $f_{\varepsilon(\tau)}(0|X)$ denotes the conditional on X density of $\varepsilon(\tau) \equiv Y^* - X'\theta(\tau)$ at 0 and

$$J_2(\tau) = E[f_{\varepsilon(\tau)}(0 \mid X)X].$$

Under Assumptions 3.2 and 4.1 it follows that

$$\sqrt{nT}\mathbb{E}_{nT}\begin{pmatrix} g_\cdot(W_{it}, \theta(\cdot)) \\ \xi_{it}\end{pmatrix} \rightsquigarrow \mathbb{G}^*(\cdot), \quad \text{in } \ell^\infty(\mathcal{T}), \tag{A.8}$$

where $\mathbb{G}^*(\cdot)$ is a zero-mean gaussian process with covariance kernel

$$\Omega(\tau, \tau') \equiv \begin{pmatrix} S(\tau, \tau') & \Omega_{g\xi}(\tau) \\ \Omega_{\xi g}(\tau') & \Omega_{\xi\xi}\end{pmatrix}, \tag{A.9}$$

and $S(\tau, \tau') \equiv (\min\{\tau, \tau'\} - \tau\tau')E(XX')$, $\Omega_{g\xi}(\tau) \equiv E[g_\tau(W, \theta(\tau))\xi]$, and $\Omega_{\xi\xi} \equiv E[\xi^2]$. The above result implies that

$$\begin{aligned}\Psi(\tau, \tau') &\equiv E[(g_\tau(W, \theta(\tau)) + J_2(\tau)\xi_{it})(g_{\tau'}(W, \theta(\tau')) + J_2(\tau')\xi_{it})'] \\ &= S(\tau, \tau') + J_2(\tau)\Omega_{\xi g}(\tau') + \Omega_{g\xi}(\tau)J_2(\tau') + J_2(\tau)\Omega_{\xi\xi}J_2(\tau')'.\end{aligned} \tag{A.10}$$

We can conclude from equation (A.7) that

$$\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot)) \rightsquigarrow \mathbb{G}(\cdot), \quad \text{in } \ell^\infty(\mathcal{T}), \tag{A.11}$$

where $\mathbb{G}(\cdot)$ is a gaussian process with covariance kernel

$$E\mathbb{G}(\tau)\mathbb{G}(\tau')' = J_1(\tau)^{-1}\Psi(\tau, \tau')[J_1(\tau')^{-1}]'. \tag{A.12}$$

□

LEMMA A.3. *If* $||v_{i,T} - v|| \to 0$ *a.s. uniformly in i as* $T \to \infty$*, and there exists a function* $q_i \geq 0$ *such that* $||v_{i,T}|| \leq q_i$ *for all i and T with* $E[\sup_i q_i] < \infty$*, then* $\sup_i E\|v_{i,T} - v\| \to 0$*, as* $T \to \infty$.

Proof: Let $h_{i,T} = ||v_{i,T} - v||$ and note that $0 \leq \sup_i h_{i,T} \leq 2\sup_i q_i$. By Fatou's Lemma

$$2E\left[\sup_i q_i\right] = E\left[\liminf_{T\to\infty}\left(2\sup_i q_i - \sup_i h_{i,T}\right)\right] \leq \liminf_{T\to\infty} E\left(2\sup_i q_i - \sup_i h_{i,T}\right), \tag{A.13}$$

$$\leq 2E\left[\sup_i q_i\right] - \limsup_{T\to\infty} E\left[\sup_i h_{i,T}\right], \tag{A.14}$$

meaning that $\limsup_{T\to\infty} E[\sup_i h_{i,T}] \leq 0$. Then, the result directly follows from

$$0 \leq \limsup_{T\to\infty} \sup_i E[h_{i,T}] \leq \limsup_{T\to\infty} E\left[\sup_i h_{i,T}\right] \leq 0. \tag{A.15}$$

□

LEMMA A.4. *Assume* $\Omega_{XX} \equiv E[(X^s_{it} - \mu^s_X)(X^s_{it} - \mu^s_X)']$ *is non-singular with finite norm,* $\frac{n}{T^a} \to 0$ *for some* $a \in (0, \infty)$ *and let Assumptions 3.2 and 4.1 hold. The within estimator of* θ_μ *satisfies Assumption 4.2 with the influence function*

$$\psi_{it} = \begin{pmatrix} \psi^0_{it} \\ \psi^s_{it} \end{pmatrix} \equiv \begin{pmatrix} Y_{it} - \mu_Y - \mu^{s\prime}_X \Omega^{-1}_{XX} \left(X^s_{it} - \mu^s_X\right) u_{it} \\ \Omega^{-1}_{XX} \left(X^s_{it} - \mu^s_X\right) u_{it} \end{pmatrix}, \tag{A.16}$$

where $X'_{it} = (1, X^{s\prime}_{it})$, $\mu^s_X \equiv E(X^s_{it})$, $\mu_Y \equiv E(Y_{it})$, u_{it} *is i.i.d. with* $E[u_{it} \mid X_i] = 0$ *and* $E[u^2_{it} \mid X_i] = X'_{it}\Omega_{UU} X_{it}$, *and* Ω_{UU} *non-singular with finite norm.*

Proof: Use the partition $\theta'_\mu = (\theta^0_\mu, \theta^{s\prime}_\mu)$, where $\theta^0_\mu \in \mathbb{R}$ and $\theta^s_\mu \in \mathbb{R}^{k-1}$. Then

$$\hat{\theta}^s_\mu \equiv \left(\mathbb{E}_{nT}\left[\tilde{X}^s_{it}\tilde{X}^{s\prime}_{it}\right]\right)^{-1} \mathbb{E}_{nT}\left[\tilde{X}^s_{it} Y_{it}\right], \text{ and } \hat{\theta}^0_\mu \equiv \mathbb{E}_{nT}(Y_{it}) - \mathbb{E}_{nT}\left(X^s_{it}\right)' \hat{\theta}^s_\mu, \tag{A.17}$$

where $\tilde{X}^s_{it} \equiv X^s_{it} - \bar{X}^s_i$ and $\bar{X}^s_i \equiv \mathbb{E}_T(X^s_{it})$. By equations (4.1) and (A.17) it follows that

$$\sqrt{nT}\left(\hat{\theta}^s_\mu - \theta^s_\mu\right) = \left(\mathbb{E}_{nT}\left[\tilde{X}^s_{it}\tilde{X}^{s\prime}_{it}\right]\right)^{-1} \sqrt{nT}\mathbb{E}_{nT}\left[\tilde{X}^s_{it} u_{it}\right], \tag{A.18}$$

where $u_{it} \equiv X'_{it}(\theta(U_{it}) - \theta_\mu)$. By Assumption 3.2, $E[u_{it}] = 0$, $E[u^2_{it} \mid X_{it}] = X'_{it}\Omega_{UU} X_{it}$, and Ω_{UU} is non-singular and has finite norm. By Assumptions 3.2 and 4.1

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}^s_{it}\tilde{X}^{s\prime}_{it} = \Omega_{XX} + o_p(1), \text{ as } n, T \to \infty, \tag{A.19}$$

for Ω_{XX} non-singular and therefore

$$\sqrt{nT}\left(\hat{\theta}^s_\mu - \theta^s_\mu\right) = \Omega^{-1}_{XX}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}^s_{it} u_{it} + o_p(1). \tag{A.20}$$

Next note that

$$\begin{aligned} \sqrt{nT}\left(\hat{\theta}^s_\mu - \theta^s_\mu\right) &= \Omega^{-1}_{XX}\left\{\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(X^s_{it} - \mu^s_X\right) u_{it} - \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\bar{X}^s_i - \mu^s_X\right) u_{it}\right\} + o_p(1), \\ &\equiv \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\psi^s_{it} - \Omega^{-1}_{XX}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\bar{X}^s_i - \mu^s_X\right) u_{it} + o_p(1), \end{aligned} \tag{A.21}$$

meaning that the result for the slope coefficients would follow provided the second term in equation (A.21) is $o_p(1)$. To show this, write this term as

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\bar{X}^s_i - \mu^s_X\right) u_{it} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\bar{X}^s_i - \mu^s_X\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varsigma_{i,T}, \tag{A.22}$$

where $\varsigma_{i,T}$ is i.i.d. across i for all T and satisfies $E[\varsigma_{i,T}] = 0$ and

$$E\|\varsigma_{i,T}\|^2 = E\left\|\left(\bar{X}^s_i - \mu^s_X\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\right\|^2 \leq E\left\|\bar{X}^s_i - \mu^s_X\right\| E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\right\| \tag{A.23}$$

$$\leq \sup_i E \left\| \bar{X}_i^s - \mu_X^s \right\| \times O(1) \to 0, \text{ as } T \to \infty, \tag{A.24}$$

since $\sup_i E\|\bar{X}_i^s - \mu_X^s\| \to 0$ as $T \to \infty$ by Lemma A.4 (note that by Assumption 4.1, the function q_i is just the upper bound of the support $\mathcal{X}$) and the fact that $\bar{X}_i^s \to^p \mu_X^s$ uniformly over i as $T \to \infty$ by similar arguments to those in Lemma A.2 provided $\frac{n}{T^a} \to 0$ for some $a \in (0, \infty)$.

Finally, from equation (A.17), $\theta_\mu^0 = \mu_Y - \mu_X^{s\prime}\theta_\mu^s$, the expansion in equation (A.21), and a few algebraic manipulations, it follows that

$$\begin{aligned}
\sqrt{nT}(\hat{\theta}_\mu^0 - \theta_\mu^0) &= \sqrt{nT}\left(\mathbb{E}_{nT}(Y_{it}) - \mu_Y + \mu_X^{s\prime}\theta_\mu^s - \mathbb{E}_{nT}\left(X_{it}^s\right)'\hat{\theta}_\mu^s\right) \\
&= \sqrt{nT}\mathbb{E}_{nT}[Y_{it} - \mu_Y] - \mu_X^{s\prime}\sqrt{nT}\left(\hat{\theta}_\mu^s - \theta_\mu^s\right) + o_p(1) \\
&= \sqrt{nT}\mathbb{E}_{nT}\left[Y_{it} - \mu_Y - \mu_X^{s\prime}\psi_{it}^s\right] + o_p(1).
\end{aligned} \tag{A.25}$$

Letting $\psi_{it}^0 \equiv Y_{it} - \mu_Y - \mu_X^{s\prime}\psi_{it}^s = Y_{it} - \mu_Y - \mu_X^{s\prime}\Omega_{XX}^{-1}(X_{it}^s - \mu_X^s)u_{it}$, the result follows. □

APPENDIX B: ESTIMATOR OF THE COVARIANCE KERNEL AND THE BOOTSTRAP

The components of the asymptotic variance in equation (A.10) can be estimated using sample analogs. The expressions below correspond to the case where $\hat{\theta}_\mu$ is the within estimator and so ψ_{it} is given by equation (A.16). They can be naturally extended to cover any other preliminary estimator satisfying Assumption 4.2.

The matrix $S(\tau, \tau')$ can be estimated by its sample counterpart

$$\hat{S}(\tau, \tau') \equiv (\min\{\tau, \tau'\} - \tau\tau')\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}X_{it}'. \tag{B.1}$$

For the matrices $J_1(\tau)$ and $J_2(\tau)$, I follow Powell (1991) and propose

$$\begin{aligned}
\hat{J}_1(\tau) &\equiv \frac{1}{2nTh_n}\sum_{i=1}^{n}\sum_{t=1}^{T} I(|\hat{\varepsilon}_{it}(\tau)| \leq h_n)X_{it}X_{it}', \\
\hat{J}_2(\tau) &\equiv \frac{1}{2nTh_n}\sum_{i=1}^{n}\sum_{t=1}^{T} I(|\hat{\varepsilon}_{it}(\tau)| \leq h_n)X_{it},
\end{aligned}$$

where $\hat{\varepsilon}_{it}(\tau) \equiv \hat{Y}_{it} - X_{it}'\hat{\theta}(\tau)$, $\hat{Y}_{it} \equiv Y_{it} - \hat{\alpha}_i$, and h_n is an appropriately chosen bandwidth such that $h_n \to 0$ and $nTh_n^2 \to \infty$. Following Koenker (2005, pp. 81 and 140), one possible choice is

$$h_n = \kappa(\Phi^{-1}(\tau + b_n) - \Phi^{-1}(\tau - b_n)), \qquad b_n = (nT)^{-1/3}z_\alpha^{2/3}\left[\frac{1.5\phi^2(\Phi^{-1}(\tau))}{2\Phi^{-1}(\tau)^2 + 1}\right]^{1/3}, \tag{B.2}$$

where κ is a robust estimate of scale, $z_\alpha = \Phi^{-1}(1 - \alpha/2)$ and α denotes the desired size of the test.

For the terms involving ξ_{it} and $g_\tau(W, \theta(\tau))$ define

$$\hat{\psi}_{it} \equiv \begin{pmatrix} Y_{it} - \hat{\mu}_Y - \hat{\mu}_X^{s\prime}\hat{\Omega}_{XX}^{-1}\tilde{X}_{it}^s\hat{u}_{it} \\ \hat{\Omega}_{XX}^{-1}\tilde{X}_{it}^s\hat{u}_{it} \end{pmatrix}, \tag{B.3}$$

where $\tilde{X}_{it}^s \equiv X_{it}^s - \bar{X}_i^s$, $\hat{u}_{it} \equiv \hat{Y}_{it} - X_{it}'\hat{\theta}_\mu$, $\hat{\mu}_Y \equiv \mathbb{E}_{nt}(Y_{it})$, $\hat{\mu}_X^s \equiv \mathbb{E}_{nT}(X_{it}^s)$ and $\hat{\Omega}_{XX} \equiv \mathbb{E}_{nT}[\tilde{X}_{it}^s\tilde{X}_{it}^{s\prime}]$. This way, we can define $\hat{\xi}_{it} \equiv \hat{\mu}_X'\hat{\psi}_{it} - \hat{u}_{it}$, where $\hat{\mu}_X \equiv \mathbb{E}_{nT}(X_{it})$. Finally, letting $\hat{g}_{\tau,it} \equiv \varphi_\tau(\hat{\varepsilon}_{it}(\tau))X_{it}$ we have the following sample counterparts for the remaining terms

$$\hat{\Omega}_{g\xi} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{g}_{\tau,it}\hat{\xi}_{it}, \quad \text{and} \quad \hat{\Omega}_{\xi\xi} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\xi}_{it}^{2}. \tag{B.4}$$

The estimator of the covariance matrix would then be $\hat{J}_1(\tau)^{-1}\hat{\Psi}(\tau,\tau')[\hat{J}_1(\tau')^{-1}]'$, where $\hat{\Psi}(\tau,\tau')$ is the matrix in equation (A.10) where all matrices have been replaced by their respective sample analogs.

In the simulations of Section 5 I use the following bootstrap algorithm to compute standard errors and confidence intervals for $\hat{\theta}(\tau)$. Let $\{Y_i^*, X_i^*\}_{i=1}^n$, $j = 1, \dots, B$, denote the jth i.i.d. sample of size n distributed according to $\hat{P}_n$, the empirical measure of $\{Y_i, X_i\}_{i=1}^n$, where $Y_i = (Y_{i1}, \dots, Y_{iT})$ and $X_i = (X_{i1}, \dots, X_{iT})$. For each $j = 1, \dots, B$ compute the two step estimator as described in Section 4 and denote this estimator by $\hat{\theta}_j^*(\tau)$. This involves computing preliminary estimators $\hat{\theta}_{\mu,j}^*$ and fixed effects $\hat{\alpha}_{i,j}^*$ for each bootstrap sample $j = 1, \dots, B$.

The bootstrap estimate of the variance covariance matrix for $\hat{\theta}(\tau)$ is given by

$$\frac{1}{B}\sum_{j=1}^{B}\left(\hat{\theta}_j^*(\tau) - \bar{\theta}^*(\tau)\right)\left(\hat{\theta}_j^*(\tau) - \bar{\theta}^*(\tau)\right)', \tag{B.5}$$

where $\bar{\theta}^*(\tau) \equiv \frac{1}{B}\sum_{j=1}^{B}\hat{\theta}_j^*(\tau)$. In the simulations of Section 5 I also report the coverage of the $1 - \alpha$ percentile interval

$$C_n^*(\tau) = \left[q_n^*(\alpha/2, \tau), q_n^*(1 - \alpha/2, \tau)\right], \tag{B.6}$$

where $q_n^*(\alpha, \tau)$ is the α-quantile of the empirical distribution of $\{\hat{\theta}_j^*(\tau)\}_{j=1}^B$. This confidence interval is translation invariant, which is a good property when working with quantile regressions. Symmetric and equally-tailed intervals can be alternatively computed using the same algorithm.