Quantile Regression for Panel Data Models with Fixed Effects and Small T: Identification and Estimation

Maria Ponomareva*

University of Western Ontario

May 8, 2011

Abstract

This paper proposes a moments-based approach to the identification and estimation of panel data quantile regression (QR) models with fixed effects when the number of time periods T is small. When the covariates have discrete support and fixed effects are pure location shifts, I show that the QR model is identified and suggest an estimator based on the recovering the distribution function from a sequence of its moments. When the covariates are continuously distributed, I show that the QR model can be identified even when fixed effects are allowed to vary across quantiles.

1 Introduction

Quantile regression (QR) models are wildly used in the empirical literature, and unlike the linear regression models based on the conditional mean restrictions, QR models allow to analyze different features of the distribution of the data while accounting for possible unobserved heterogeneity. Identification and estimation of linear QR models in the cross-sectional case is barely harder than the identification and estimation of a linear conditional mean model,

*I am very grateful to Jason Abrevaya, Ivan Canay, Wenxin Jiang and Elie Tamer for valuable comments and discussions. Comments and suggestions are welcome. All remaining errors are mine. E-mail: mponomar@uwo.ca

and the asymptotic theory for QR estimators is well developed for cross-sectional data. However, when QR model is combined with panel data model that controls for the unobserved heterogeneity that is constant over the time via fixed effects, identification and estimation become really complicated. This happens because the standard methods that difference out fixed effects are no longer applicable since the quantile of the difference in general are not equal to the difference in quantiles but rather are some intractable object. The problem becomes even more complicated when the number of time periods is small and therefore we cannot directly estimate the unobserved fixed effects.

In this paper I provide two sets of sufficient conditions that allow to identify QR panel data models with fixed effects under different assumption about the distribution of fixed effects. In particular, I treat separately two cases: when fixed effects represent pure location shifts and when fixed effects are allowed to vary with the quantile. For each of this cases I present the condition under which the marginal quantile effects are identified when the number of periods is fixed. In the case when fixed effects are pure location shifts, I propose the estimation procedure that is based on the recovery of the distribution function from the sequence of the consistent estimators of its moments.

The majority of the literature that studies QR models for panel data with fixed effects propose inference procedures based on the assumption that the number of periods T goes to infinity when the sample size n goes to infinity. This assumption allows to estimate unobservable fixed effects α_i. Under this assumption, Koenker (2004) and Lamarche (2010) suggest a penalized quantile regression estimator that simultaneously estimates quantile regression coefficients for a set of quantiles $\{0 < \tau_1 < \ldots < \tau_m\}$ and fixed effects. Galvao (2008) adopts a similar approach in the context of dynamic panel data. Canay (2010) introduces a different approach that does not require specifying a penalty parameter. He suggests a simple two-step procedure that relies on the transformation of the data and where the unobserved fixed effects are estimated at the first step. Koenker (2004), Lamarche (2010) and Canay (2010) assume that fixed effects α_i have a pure locations shift effect, while Galvao (2008) allows fixed effect to depend upon the quantile of interest.

When the number of periods T is small, one cannot estimate fixed effects consistently any longer. Abrevaya and Dahl (2008) impose a particular structure on the relationship between unobserved fixed effects and regressors and quantiles. As a result they obtain a correlated random coefficients model that can be estimated consistently using standard quantile regression technique. Rosen (2009) focuses on the identification of a quantile regression coefficients for a single conditional quantile restriction rather than for the whole set of quantiles $0 < \tau < 1$. He imposes no restrictions on the distribution of fixed effects and shows that under rather weak assumptions linear conditional quantile function can be at least partially identified and provides sufficient conditions for point identification.

In this paper I treat the QR panel data model as a special case of a random coefficients model. Related papers that study random coefficients model in the context of panel data include Graham and Powell (2008) and Graham, Hahn, and Powell (2009). A recent paper by Arellano and Bonhomme (2009) focuses on the identification and estimation of certain features of the distribution of random coefficients in panel data models, including first and second moments of those distributions.

The rest of this paper is organized as follows. Section 2 presents the model and outlines identification and moments-based estimation strategy that I propose in this paper. In section 3, I present a set of assumptions sufficient for the identification of QR panel data model when the covariates have discrete distribution and fixed effects are pure location shifts. I propose an estimator based on the sequence of moment estimators, where each of moment estimators is $n^{1/2}$-consistent and asymptotically normally distributed in the case of discrete covariates. In section 4, I consider the case when the regressors have continuous distribution and give a set of sufficient conditions that allows to identify the QR model even when fixed effects are allowed to depend on quantile. Here I also discuss the possibility of partial identification of QR panel data model with discrete regressors when fixed effects are allowed to vary with quantile. Finally, section 5 concludes. All proofs of the results are collected in the Appendix.

2 The Model

I consider the following representation of quantile regression model:

$$Y_{it} = X_{it}'\theta(U_{it}) + \alpha_i, \text{ where } U_{it}|(X_i, \alpha_i) \sim U[0,1],\ i = 1, \dots, n,\ t = 1, 2. \tag{1}$$

Here $X_i = (X'_{i1}, X'_{i2})'$ is a random vector of regressors and the function $\tau \to X'_{it}\theta(\tau)$ is assumed to be strictly increasing on the interval $(0, 1)$ for any given realization of X_{it}.[1] This is a convenient representation of quantile regression model that is due to Doksum (1974). The data available to a researcher consists of observations $\{(Y_{it}, X_{it}), i = 1, \ldots, n;\ t = 1, 2\}$. Fixed effects α_i's are not observable and can be arbitrary related to random vectors X_i and (U_{i1}, U_{i2}). The object of interest in this model is the vector function $\theta(\cdot)$. In particular, a researcher might be interested in a set of values of $\theta(\tau_j)$ evaluated at a number of quantiles $\{\tau_1, \ldots, \tau_k\}$. The condition that $U_{it} \perp \alpha_i$ which is implicitly implied in model (1) will be removed later in section 4.

For a fixed quantile $0 < \tau < 1$, this model represents a special case of panel quantile regression analyzed in Rosen (2009). In particular, let $\beta = \theta(\tau)$ and define $\tilde{U}_{it}(\tau) = X'_{it}(\theta(U_{it}) - \theta(\tau))$. Denote the conditional τ-quantile of $\tilde{U}_{it}(\tau)$ by $Q_{\tilde{U}_{it}(\tau)}(\tau|X_i)$. Then for this fixed value of τ the model in (1) can be written as

$$Y_{it} = X'_{it}\beta + \alpha_i + \tilde{U}_{it}, \text{ where } Q_{\tilde{U}_{it}(\tau)}(\tau|X_i) = 0. \tag{2}$$

Rosen shows that it is impossible to place any meaningful restrictions on the parameter β in (2) without imposing any additional assumptions about the behavior of $\tilde{U}_{it}|X_i$ besides the conditional quantile restriction. He shows then that a sufficient condition for at least weak identification of β is that events $\{\tilde{U}_{i1}(\tau) < 0\}$ and $\{\tilde{U}_{i2}(\tau) < 0\}$ are independent conditional on X_i. Note that if U_{i1} and U_{i2} are independent conditional on X_i, then for any $0 < \tau < 1$ the events $\{\tilde{U}_{i1}(\tau) < 0\}$ and $\{\tilde{U}_{i2}(\tau) < 0\}$ are also independent conditional on X_i, so that for any τ, $\theta(\tau)$ is at least partially identified. The assumption that U_{i1} and U_{i2} are independent conditional on X_i is also one of the key identifying assumptions in Canay (2010). Another key identifying assumption in Canay (2010) is that α_i has a pure location shift effect, i.e. α_i is independent of U_{i1} and U_{i2} conditional on X_i. The approach proposed by Rosen (2009) for a fixed quantile does not require to make such an assumption, however in this case point identification of $\theta(\tau)$ requires that at least one of the covariates is continuously distributed.

[1]Throughout the paper I use upper case letters to denote random variables or the element of the random sample, and lower case letters to represent a particular realization or the point in the support of the corresponding random variable.

The main contribution of the present paper is that it does not require the number of periods, T, to grow together with the sample size and offers a relatively simple way to estimate the parameter of interest $\theta(\tau)$ that does not require deconvolution. In particular, in the case when all covariates are discrete and α_i has a pure location shift effect, this paper proposes an estimator that is based on the recovering of the distribution from its moments. However, if at least $(d-1)$ covariates are continuous, I show that then under certain constraints it is possible to identify $\theta(\cdot)$ even if α_i is allowed to depend on U_{i1} and U_{i2} and propose an moments-based estimator for this case. Next paragraph briefly outlines the general inference procedure proposed in this paper.

Outline of the Inference Procedure: Identification and estimation of the function $\tau \to x_{it}'\theta(\tau)$ for $0 < \tau < 1$ essentially amounts to identification and estimation of the distribution of $X_{it}'\theta(U_{it})$ conditional on X_{it}. Once we obtain consistent estimators for those conditional distribution functions, the inference procedure becomes really simple: we can sample from those distribution and estimate $\theta(\tau)$ for any given $0 < \tau < 1$ by an ordinary quantile regression. The standard error of such an estimator based on a sampling will depend only on the standard error of the estimators of the conditional distribution function. One of the many ways to estimate a distribution function is to estimate its moments and then if the distribution is uniquely defined by its moments, we can recover the distribution function from those moments. This approach is used e.g. in Beran and Hall (1992) to estimate distributions in a certain class of random coefficients regression models. Finally, the particular representation of the model in (1) suggests that in the case of continuous covariates one can use a sequence of OLS estimators to estimate those moments. The estimator is even simpler when the vector of covariates has finite support.

As it was already mentioned above, the set of key identifying assumptions is different depending on whether the covariates have finite or continuous support (in the second case we can relax the assumption that fixed effects are pure location shifts). Therefore, the next two sections treat each case separately, carefully summarizing the set of identifying restrictions in each case and proposing an estimator that is consistent under the corresponding set of assumptions.

3 Identification and Estimation with Discrete Covariates

In many empirical applications the regressors (such as age, gender, education level etc.) have a discrete joint distribution. In this section I provide Throughout the discussion in this section, I assume that random vector $X_i = (X_{i1}', X_{i2}')')$ has a discrete distribution with finite support $\mathcal{X} = suppX_i$ and let $x = (x_1', x_2')'$ denote a typical element of this set. Let $d = dimX_{it}$. For any given $x \in \mathcal{X}$ and any $t = 1, 2$ define

$$\mu_k(x_t) = E[(X_{it}'\theta(U_{it}))^k | X_{it} = x_t].$$

Here $\mu_k(x_t)$ is the k^{th} order moment of a scalar random variable $X_{it}'\theta(U_{it})$.

Assumption 1. *For any $x \in \mathcal{X}$, distributions of random variables $X_{it}'\theta(U_{it})|X_{it} = x_t$ for $t = 1, 2$ and distribution of $\alpha_i|X_i = x$ are uniquely determined by its moments (all assumed to be finite).*

Assumption 1 implies that if the moments of the corresponding distributions are identified, then we can identify the distribution itself. Assumption 1 holds if, for example, for any $1 \leq m \leq d$, the function $\theta_d(\cdot)$ is bounded on $[0, 1]$.

Assumption 2. *Conditional on X_i, $U_{it} \perp \alpha_i$ and $U_{it} \sim U[0, 1]$ for $t = 1, 2$. Also, U_{i1} and U_{i2} are independent.*

Assumption 2 is a key identifying assumption used in Canay (2010). It rules out the case when α_i may depend on U_{i1} and U_{i2}, so that α_i is a pure location shift effect (that is, the same for all quantiles). The following Theorem provides the identification result under those assumptions that will be used to construct an estimator for $\theta(\tau)$ based on a consistent estimators of the moments of $X_{it}'\theta(U_{it})|X_{it} = x_t$.

Theorem 1. *Suppose that $X_i = (X_{i1}', X_{i2}')'$ has a finite support $\mathcal{X}$ and that Assumptions 1 and 2 are satisfied by the distributions of $X_{i1}'\theta(U_{i1})$, $X_{i2}'\theta(U_{i2})$ and α_i. Additionally, suppose that the matrix $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$ has full rank. Then for any integer $1 \leq k < +\infty$ and any $x \in \mathcal{X}$, $\mu_k(x_t)$ and $E[\alpha_i^k | X_i = x]$ are identified.*

Remark 3.1. Theorem 1 together with Assumption 1 imply that the distributions of $X_{it}'\theta(U_{it})$ and α_i conditional on X_i are identified. Also, this Theorem provides an alternative proof of the identification result for short panel in Canay (2010).

Remark 3.2. The result in Theorem 1 can be extended to the case with more than two periods, so that with the appropriate modification of Assumptions 1 and 2, the identification claim is also true in the case $t = 1, \ldots, T$, where $T \geq 2$.

Remark 3.3. Assumption that $X_i = (X_{i1}', X_{i2}')'$ has a finite support is not essential for the identification argument. If some of the regressors are continuously distributed and if Assumptions 1 and 2 hold on the support of X_i, then one can still identify the sequence of moments $\{\mu_j(x_t), E[\alpha_i^j|X_i = x], 1 \leq j \leq k\}$ for any $1 \leq k \leq \infty$.

Estimation. Theorem 1 implies that if we can estimate a sequence of moments $\{\mu_k(x_t), k = 1, \ldots\}$ for any x in the support of X_i, then we can recover the distribution of a scalar random variable $x_t'\theta(U_{it})$. The statistic literature offers a variety of techniques to estimate the distribution from its moments. Beran and Hall (1992) have a review some of these methods, including approximations based on series and discrete approximations.[2] I start by showing how for any given $k \in \mathbb{N}$ one can estimate moments $\{\mu_j(x_t), 1 \leq j \leq k\}$ sequentially. The by-product of this procedure are estimators for $\{E[\alpha_i^j|X_i = x], 1 \leq j \leq k\}$.

Recall that the support of X_i is finite. First, note that given Assumption **??**, for any $x = (x_1, x_2) \in \mathcal{X}$ we have

$$E[Y_{i2} - Y_{i1}|X_i = x] = (x_2 - x_1)'\theta_\mu, \tag{3}$$

where $\theta_\mu = E[\theta(U_{it})]$. Thus, we can compute a $n^{1/2}$-consistent estimators of $\mu_k(x_1) = x_1'\theta_\mu$ and $\mu_k(x_2) = x_2'\theta_\mu$ from a simple linear regression of the difference $Y_{i2} - Y_{i1}$ on $X_{i2} - X_{i1}$. In particular, let $\hat{\theta}_\mu$ be an OLS estimator of θ_μ in (**??**) and let $\hat{\mu}_1(x_t) = x_t'\hat{\theta}_\mu$. We can also

[2]The techniques discussed in Beran and Hall (1992) rely on a certain uniform approximation result for the sequence of moments. Below I will provide a corresponding uniform approximation result for the sequence of moments estimators proposed in this paper. Given this uniform approximation result, the discussion of the distribution recovery techniques for the particular case of panel data model closely follows the discussion by Beran and Hall and therefore is omitted here.

estimate $E[\alpha_i|X_i = x]$ $n^{1/2}$-consistently by e.g.

$$\hat{E}[\alpha_i|X_i = x] = \frac{1}{2}\left(\hat{E}[Y_{i1} + Y_{i2}|X_i = x] - (\hat{\mu}_1(x_1) + \hat{\mu}_1(x_2))\right). \tag{4}$$

Here we use the following conventional notation: for any random variable V_i define

$$\hat{E}[V_i|X_i = x] = \frac{\sum_{i=1}^{n} V_i 1\{X_i = x\}}{\sum_{i=1}^{n} 1\{X_i = x\}},$$

where $1\{\cdot\}$ is the indicator function.

For a moment, assume that it is known that $\theta_\mu = 0$. Suppose now that the estimates $\{(\hat{\mu}_j(x_t), \hat{E}[\alpha_i^j|X_i = x]), 1 \le j \le k-1\}$ has already been computed. Then we can construct estimators $\hat{\mu}_k(x_t)$ and $\hat{E}[\alpha_i^k|X_i = x]$ as (see the proof of Theorem 1 in the Appendix):

$$\begin{aligned} \hat{E}[\alpha_i^k|X_i = x] &= \hat{E}[Y_{i1}^{k-1}Y_{i2}|X_i = x] - \sum_{j=0}^{k-2}\binom{k-1}{j}\hat{\mu}_{k-1-j}(x_1)\hat{E}[\alpha_i^{j+1}|X_i = x] \\ \hat{\mu}_k(x_t) &= \hat{E}[Y_{it}^k|X_i = x] - \sum_{j=1}^{k}\binom{k}{j}\hat{\mu}_{k-j}(x_t)\hat{E}[\alpha_i^j|X_i = x] \end{aligned} \tag{5}$$

For each fixed k, the estimators $\hat{\mu}_k(x_t)$ and $\hat{E}[\alpha_i^k|X_i = x]$ defined either in (3) and (4) for $k = 1$ or in (5) for $k > 1$ are $n^{1/2}$ consistent and asymptotically normally distributed when the data are i.i.d. sample and if certain moments of the distributions of random variables $X_{it}'\theta(U_{it})$ and α_i exist and are finite. In particular, suppose that the following assumptions are satisfied:

Assumption 3. *(i) The data $(Y_{i1}, Yi2, X_{i1}, X_{i2}, \alpha_i)$ are i.i.d. sample from $(\Omega, \mathcal{F}, P)$.*

(ii) The support of $X_i = (X_{i1}', X_{i2}')'$ is finite. For any given $x = (x_1, x_2) \in \mathcal{X} = suppX_i$, the distributions of random variables $x_t'\theta(U_{it})$ and $\alpha_i|X_i = x$ are essentially bounded. Also, $E[\theta(U_{it})] = 0$.

(iii) The $d \times d$ matrix $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$ has full rank.

Here Assumption 3(i) is a standard random sampling condition. Assumption 3(ii) guarantees that for any point in the support of X_i, the distributions of $x_t'\theta(U_{it})$ and $\alpha_i|X_i = x$ are uniquely determined by their moments and that all those moments are finite. Finally, Assumption 3(iii) is a standard full rank condition that is required to identify $\theta_\mu = E[\theta(U_{it})]$. The Theorem below shows that if number of moments to be estimated, k, goes to infinity at a certain rate, then $\hat{\mu}_k(x_t)$ and $\hat{E}[\alpha_i^k|X_i = x]$ converge *uniformly* to $\mu_k(x_t)$ and $E[\alpha_i^k|X_i = x]$, respectively.

Theorem 2. *Suppose that Assumptions 1, 2 and 3 are satisfied. Then for any $\delta > 0$ there exists $\eta > 0$ such that with probability 1,*

$$\max_{1 \le k \le (\eta \log n)^{1/2}} \left(|\hat{\mu}_k(x_t) - \mu_k(x_t)| + |\hat{E}[\alpha^k|x] - E[\alpha^k|x]| \right) = O(n^{-1/2+\delta}) \text{ as } n \to \infty.$$

Remark 3.4. The requirement that $E[\theta(U_{it})] = 0$ can be easily relaxed. In this case the result in Theorem 2 carries through if instead of observable $(Y_{it}, X_{it}, \alpha_i)$ we consider partially observable $(Y_{it}^*, X_{it}, \alpha_i)$ where $Y_{it}^* = Y_{it} - X_{it}'\hat{\theta}_\mu$ and use $\hat{Y}_{it}^* = Y_{it} - X_{it}'\hat{\theta}_\mu$ as an estimator for the unobservable Y_{it}^*. Assumption 3(ii) that the conditional distributions are essentially bounded allows us to do this. Estimation procedure in (5) must be adjusted accordingly.

Remark 3.5. When some of the components of X_{it} are continuously distributed, we cannot estimate $\hat{\mu}_k(x_t)$ and $\hat{E}[\alpha_i^k|X_i = x]$ $n^{1/2}$-consistently any longer without imposing strong parametric assumptions about the distribution of α_i conditional on X_i. However, one can use any nonparametric methods of conditional moment estimation. In this case, the rate of convergence will be slower and the corresponding rate in Theorem **??** must be adjusted accordingly.

Remark 3.6. Theorem 2 allows us to estimate the distributions of both $x_t'\theta(U_{it})$ and $\alpha_i|X_i = x$ for any given $x \in \mathcal{X}$ from a finite sequence of moments.[3] There is no general rule on how many moment to choose for this estimation for a given sample size. However, condition that $\eta = O(\delta)$ as $\delta \to 0$ (see the proof of Theorem 2 in the Appendix) suggests that one should choose k much smaller than the sample size n.

Suppose now that we have estimated conditional distributions based on the estimators

[3]For the review of some methods see Beran and Hall (1992). Also, see Greaves (1982), Mnatsakanov and Hakobyan (2009).

of the first k moments. In particular, define $S_{it} = X'_{it}\theta(U_{it})$ and let $\hat{F}_t(\cdot|X_i = x)$ be the estimator of $F_t(\cdot|X_i = x)$ - the conditional distribution function of $S_{it} = X'_{it}\theta(U_{it})$. An easy way to estimate conditional quantiles of S_{it} conditional on X_i is to draw a random number from the distributions $\hat{F}_t(\cdot|X_i = x)$ for each X_{it} in the original sample. This way we get a random sample $\{\hat{S}_{it}, X_{it}, i = 1, \ldots, n, t = 1, 2\}$ which can be used to estimate $\theta(\tau)$ for any given $0 < \tau < 1$. Suppose that for any $0 < \tau < 1$, $\theta(\tau)$ belongs to a compact set Θ.[4] Then we can estimate $\theta(\tau)$ with

$$\hat{\theta}(\tau) = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(\hat{S}_{i1} - X'_{i1}\theta\right), \tag{6}$$

where $\rho_\tau(u) = u(\tau - 1\{u < 0\})$. To estimate standard error of the estimator in (6) one can use a nonparametric bootstrap. In particular, R bootstrap samples of size n can be drawn from the data, and bootstrap distribution of $\hat{\theta}(\tau)$ can be computed based on the bootstrap sample $\{\hat{\theta}^*_r(\tau), r = 1, \ldots, R\}$.

The discussion in the present section relies on the assumption that random vector $X_i = (X'_{i1}, X'_{i2})'$ has finite support. Next section relaxes this assumption and shows how one can use a continuous support to identify vector function $\theta(\cdot)$ without imposing the restrictions that fixed effects α_i's are pure location shifts.

4 Identification and Estimation with Continuous Covariates

When at least $(d - 1)$ of the covariates are continuous, it is possible to identify function $\tau \to x'_t\theta(\tau)$ without imposing the restriction that fixed effects α_i's are independent of $U_i = (U_{i1}, Ui2)$ conditional on X_i. As before, for any point $x \in \mathcal{X} = suppX_i$ we define

$$\mu_k(x_t) = E[(X'_{it}\theta(U_{it}))^k \,|X_{it} = x_t].$$

[4]The existence of such compact set Θ is implied by Assumption 3(ii).

Set $\mu_0(x_t) = 1$. Since $E[Y_{i2} - Yi1|X_i = x] = (x_2 - x_1)\theta_\mu$, then if the matrix $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$ has full rank, same arguments can be used to show that $\mu_1(x_t)$ is identified for any $x = (x_1, x_2) \in \mathcal{X}$. In particular,

$$\mu_1(x_t) = x_t\theta_\mu.$$

Note that $\mu_1(x_t)$ is a multivariate polynomial of degree 1 in the elements of vector x_t. In fact, any $\mu_k(x_t)$ has a similar structure: for any integer k, $\mu_k(x_t)$ is a homogeneous multivariate polynomial of degree k in the elements of vector x_t. That is, we have the following expression for the k^{th} order conditional moment of random variable $X_{it}'\theta(U_{it})$

$$\mu_k(x_t) = \sum_{l_1+\ldots+l_d=k} c_{l_1,\ldots,l_d}(k)x_{t,1}^{l_1} \cdot \ldots \cdot x_{t,d}^{l_d}, \tag{7}$$

where $d = \dim(X_{it})$. Consider the following assumption.

Assumption 4. *Conditional on X_i, $U_{it} \sim U[0,1]$ for $t = 1, 2$. Also, U_{i1} and U_{i2} are independent.*

Unlike Assumption 2 employed in the previous section, Assumption 4 does not require fixed effects α_i's to be independent of $U_i = (U_{i1}, U_{i2})$ conditional on X_i. The proof of the following identification result is given in the Appendix.

Theorem 3. *Suppose that at least $(d-1)$ of the components of random vector $X_i = (X_{i1}', X_{i2}')'$ are continuously distributed. Also, let Assumptions 1 and 4 be satisfied by the distributions of $X_{i1}'\theta(U_{i1})$, $X_{i2}'\theta(U_{i2})$ and α_i. Additionally, suppose that the matrix $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$ has full rank. Then for any integer $1 \leq k < +\infty$ and any $x \in \mathcal{X}$, $\mu_k(x_t)$ is identified.*

Remark 4.1. Theorem 1 together with Assumption 1 imply that the distributions of $X_{it}'\theta(U_{it})$ and α_i conditional on X_i are identified.

Theorem 1 suggests that one can estimate a sequence of moments $\{\mu_j(x_t), 1 \leq j \leq k\}$ by a sequence of simple linear regressions. Note, however, that unlike the estimation procedure discussed in the previous section, the k^{th} step requires to estimate a linear regression model whose dimension is $\binom{d+k-1}{d-1} = O(k^d)$ as $k \to \infty$. Therefore, given the dimension of the problem $d = \dim(X_{it})$, the number of moments to be estimated must be small relative to the size of the sample, n. In general, the following sequential procedure can be used to estimate the sequence of moments $\{\mu_j(x_t), 1 \leq j \leq k\}$ for any $x \in \mathcal{X}$:

1. Let $\mathbf{X}_t = (X'_{it})$ be a $n \times d$ matrix of covariates and $\mathbf{Y}_t = (Y_{it})$ be a $n \times 1$ vector of dependent variables for period $t = 1, 2$. Then

$$\hat{\theta}_\mu = ((\mathbf{X}_2 - \mathbf{X}_1)'(\mathbf{X}_2 - \mathbf{X}_1))^{-1} ((\mathbf{X}_2 - \mathbf{X}_1)'(\mathbf{Y}_2 - \mathbf{Y}_1))$$

   is the OLS estimator of θ_μ. For any $x \in \mathcal{X}$ define $\hat{\mu}_1(x_t) = x'_t\hat{\theta}_\mu$.

2. Suppose that we have a sequence of $n^{1/2}$-consistent estimators $\{\hat{\mu}_j(x_t), 1 \leq j \leq k-1\}$ for any $x \in \mathcal{X}$. Define

$$\hat{W}_i = (Y_{i2} - Y_{i1})^k - \sum_{j=1}^{k-1} \binom{k}{j} (-1)^j \hat{\mu}_{k-j}(X_{i1})\hat{\mu}_j(X_{i2}) \tag{8}$$

   and let $Z_i(k)$ be a $\binom{d+k-1}{d-1} \times 1$ vector with a typical element $X_{i1,1}^{l_1} \ldots X_{i1,d}^{l_d} + (-1)^k X_{i2,1}^{l_1} \ldots X_{i2,d}^{l_d}$. Finally, let $\beta(k)$ be a $\binom{d+k-1}{d-1} \times 1$ vector of parameters with a typical element $c_{l_1,\ldots,l_d}(k)$.

3. Define $\hat{\beta}(k) = ((\mathbf{Z}(k))'(\mathbf{Z}(k)))^{-1} ((\mathbf{Z}(k))'\mathbf{W})$. Then

$$\hat{\mu}_k(x_t) = z_t(k)'\hat{\beta},$$

   where a typical element of the $\binom{d+k-1}{d-1} \times 1$ vector $z_t(k)$ is $x_{1,1}^{l_1} \ldots x_{1,d}^{l_d}$.

Remark 4.2. This procedure allows us to estimate only a finite number of moments given the fixed sample size n. This was not the case in the previous section, where for a given sample size we are able to estimate as many moments as we want. This is essentially the same issue that arises in estimating conditional moments by series methods, where the number of terms in a series approximation cannot be made arbitrary big given the sample size.[5]

[5] See e.g. Ullah and Pagan (1999) for an overview of series estimator for nonparametric conditional moments problems.

Remark 4.3. For each fixed k, the estimators $\hat{\mu}_k(x_t)$ defined above are $n^{1/2}$-consistent and asymptotically normally distributed when the data are i.i.d. sample and when certain moments of the distributions of random variables $X_{it}'\theta(U_{it})$ exist and are finite.

Remark 4.4. When the number of periods, T, is large, the upper bound on the number of moment one can estimate for a fixed sample size becomes higher. Therefore, if $T \to \infty$ as $n \to \infty$, then (i) moments can be estimated $(nT)^{1/2}$-consistently; (ii) as the result, the error of the estimator of the distributions of $X_{it}'\theta(U_{it})$ based on those moments becomes smaller.

Once we estimated a sequence of moments $\{\hat{\mu}_j(x_t), 1 \leq j \leq k\}$, an inference procedure can be based on the same bootstrap approximation method discussed in the previous section. That is, we can sample from the estimate of the distribution of $X_{it}\theta(U_{it})$ conditional on X_{it}, obtain the estimator $\hat{\theta}(\tau)$ in a same way, and then employ a nonparametric bootstrap to estimate standard error of $\hat{\theta}(\tau)$.

Partial Identification with a Fixed Number of Moments: Remark 4.3 suggests that if we e.g. fix $k = 4$, then we can estimate first 4 moments $n^{1/2}$-consistently. To recover the distribution of $x_t\theta(U_{it})$ one needs to identify all moments of this distribution. However, it is possible to *partially* identify the distribution of $x_t\theta(U_{it})$ from the first k moments. That is, given e.g. first 4 moments one can construct upper and lower bounds on $F_t(\cdot|x_t)$ - the distribution function of random variable $S_{it}(x_t) = x_t'\theta(U_{it})$. Lasserre (2002) offers a method that allows to estimate those bounds based on the first k moments. Bounds on the distribution $F_t(\cdot|x_t)$ in turn imply bounds on the QR coefficients $\theta(\tau)$. Note that even when X_{it} has a discrete support, we can still identify first k moments of the distribution of $S_{it}(x_t)$ from the moments of the difference $Y_{i2} - Y_{i1}|X_i = x$ if the number of the support points is high enough. That is, it is possible to impose meaningful restrictions on the QR coefficients $\theta(\tau)$ even when the covariates have discrete distribution and when fixed effects α_i's are allowed to depend on $U_i = (U_{i1}, U_{i2})$, i.e. when we allow fixed effects to be different for different quantiles.

5 Conclusion

This paper offers a novel approach to the identification and estimation of the linear quantile regression panel data models with fixed effects when T is small. This approach is based on the identification and estimation of moments of the conditional distribution of $X_{it}\theta(U_{it})$ conditional on X_{it}. In particular, I show that when regressors are discrete and if fixed effects α_i are independent of the quantiles, then the QR model is fully identified including the conditional distribution of fixed effects and that the moments of those conditional distributions can be estimated at a parametric rate. I provide a uniform convergence result for the sequence of the estimators of those moments that allows us to estimate the conditional distribution of $X_{it}\theta(U_{it})$ and therefore its quantiles $X_{it}\theta(\tau)$ for any $0 < \tau < 1$. Finally, I show that when the covariates are continuously distributed, the linear quantile regression model is identified even when fixed effects α_i are allowed to vary across quantiles. The identification results in this paper are based on the identification of the sequence of moments of the corresponding distributions and therefore are valid under the assumption that those distributions are uniquely defined by their moments. Finally, I suggest an approach that can partially identify the parameters of interest given only a fixed number of moments.

One of the limitations of the present paper is that no asymptotic results are available for the estimators proposed here. Therefore, future research should include checking the validity of the conjecture in this paper that the nonparametric bootstrap can be used to estimate standard errors of the estimator proposed here. Also the uniform convergency of the sequence of moments estimators in the case when covariates are continuously distributed remains an open question. In particular, it seems plausible that the convergence rate will be slower than in the case with discrete covariates.

Appendix

A Proofs

Proof of Theorem 1: For the ease of presentation, index i is omitted here. Let $\mathcal{X} = suppX_i$ and consider any $x = (x_1, x_2) \in \mathcal{X}$. Recall that by Assumption 2, $U_t \perp \alpha$ conditional on $X =$

x. Note that $\mu_1(x_1) = x_1\theta_\mu$ and similarly $\mu_1(x_2) = x_2\theta_\mu$. Therefore, $\mu_1(x_1)$ is identified from $E[Y_2 - Y_1|X = (x_1, x_2)] = (x_2 - x_1)'\theta_\mu$ when the matrix $E[(X_2 - X_1)(X_2 - X_1)']$ has full rank. This implies that $E[\alpha|X = (x_1, x_2)] = \frac{1}{2}(E[Y_1|X = x] + E[Y_2|X = x] - \mu_1(x_1) - \mu_1(x_2))$ is also identified. For the ease of the following presentation, assume without loss of generality that $\theta_\mu = 0$. Define $\mu_0(x_t) = 1$ and $E[\alpha^0|X = x] = 1$.

Suppose that for any $j \in \mathbb{N}$, $1 \leq j \leq k-1$ both $E[\alpha^j|X = x]$ and $\mu_j(x_t)$ are identified. I will show that in this case $E[\alpha^k|X = x]$ and $\mu_k(x_t)$ are also identified. In order to do this, consider $E[Y_1^{k-1}Y_2|X = x] = E[Y_1^{k-1}(x_2\theta(U_2) + \alpha)|X = x] = E[y_1^{k-1}\alpha|x]$ since U_1 and U_2 are independent conditional on x and $E[X_2\theta(U_2)|X = x] = 0$. Note that Y_1^{k-1} is a polynomial of degree $k-1$ in α and $X_1'\theta(U_1)$. In particular, $Y_1^{k-1} = \sum_{j=0}^{k-1} \binom{k-1}{j}(X_1'\theta(U_1))^{k-1-j}\alpha^j$. Since $U_1 \perp \alpha$ conditional on X, finally have $E[Y_1^{k-1}Y_2|X = x] = \sum_{j=0}^{k-1}\binom{k-1}{j}\mu_{k-1-j}(x_1)E[\alpha^{j+1}|X = x]$, so that $E[\alpha^k|X = x] = E[Y_1^{k-1}Y_2|X = x] - \sum_{j=0}^{k-2}\binom{k-1}{j}\mu_{k-1-j}(x_1)E[\alpha^{j+1}|X = x]$ and therefore $E[\alpha^k|X = x]$ is identified if $E[\alpha^j|X = x]$ and $\mu_j(x_t)$ are identified for any $1 \leq j \leq k-1$. Similarly, $E[Y_t^k|X = x] = \sum_{j=0}^{k}\binom{k}{j}\mu_{k-j}(x_t)E[\alpha^j|X = x]$, so that $\mu_k(x_t) = E[Y_t^k|X = x] - \sum_{j=1}^{k}\binom{k}{j}\mu_{k-j}(x_t)E[\alpha^j|X = x]$ is identified if $E[\alpha^j|X = x]$ and $\mu_j(x_t)$ are identified for any $1 \leq j \leq k-1$. $\square$

Proof of Theorem 2: The majority of the proof follows the proof of the similar result in Beran and Hall (1992). Let $x = (x_1, x_2)$ be any point in $\mathcal{X} = supp(X)$. By Assumption 3(ii), $\mathcal{X}$ is a finite set. Therefore, it is sufficient to show that the claim is true for a given $x \in \mathcal{X}$. Recall that

$$\begin{aligned}
\hat{E}[\alpha^k|X = x] &= \hat{E}[Y_1^{k-1}Y_2|X = x] - \sum_{j=0}^{k-2}\binom{k-1}{j}\hat{\mu}_{k-1-j}(x_1)\hat{E}[\alpha^{j+1}|X = x], \\
\hat{\mu}_k(x_t) &= \hat{E}[Y_t^k|X = x] - \sum_{j=1}^{k}\binom{k}{j}\hat{\mu}_{k-j}(x_t)\hat{E}[\alpha^j|X = x].
\end{aligned}$$

Define the following set

$$A_{k,1}(x) = \left\{ |\hat{\mu}_j(x_t) - \mu_j(x_t)| + |\hat{E}[\alpha^j|X=x] - E[\alpha^j|X=x]| \le (n^{-1}\log n)^{1/2} C^{\sum_{l=1}^{j} l},\ 1 \le j \le k-1 \right\} \tag{9}$$

We have $\mu_k(x_t) = E[Y_t^k|X=x] - \sum_{j=1}^{k} \binom{k}{j} \mu_{k-j}(x_t) E[\alpha^j|X=x]$, so that

$$\hat{\mu}_k(x_t) + \hat{E}[\alpha^k|X=x] = \hat{E}[Y_t^k|X=x] - \sum_{j=1}^{k-1} \binom{k}{j} \hat{\mu}_{k-j}(x_t) \hat{E}[\alpha^j|X=x],$$

and therefore

$$\begin{aligned} &|\hat{\mu}_k(x_t) - \mu_k(x_t)| + |\hat{E}[\alpha^k|X=x] - E[\alpha^k|X=x]| \le |\hat{E}[Y_t^k|X=x] - E[Y_t^k|X=x]| \\ &+ \sum_{j=1}^{k-1} \binom{k}{j} |\hat{\mu}_{k-j}(x_t)\hat{E}[\alpha^j|X=x] - \mu_{k-j}(x_t) E[\alpha^j|X=x]| \end{aligned} \tag{10}$$

Let $M > 1$ denote the upper bound on each of $\operatorname{ess\,sup} |x_t' \theta U_t|$ and $\operatorname{ess\,sup} |\alpha|$ conditional on $X_i = x$ for any $x \in \mathcal{X}$. By Bernstein's inequality, for any s we have

$$P\{|\hat{E}[Y_t^k|X=x] - E[Y_t^k|X=x]| \ge n^{-1/2} M^k s\} \le 2e^{-s^2/4} \tag{11}$$

Note that the second term of the right-hand side of (10) is bounded by:

$$\begin{aligned} &\sum_{j=1}^{k-1} \binom{k}{j} |\hat{\mu}_{k-j}(x_t)\hat{E}[\alpha^j|X=x] - \mu_{k-j}(x_t)E[\alpha^j|X=x]| \\ &\le M^k \sum_{j=1}^{k-1} \binom{k}{j} \left(|\hat{\mu}_j(x_t) - \mu_j(x_t)| + |\hat{E}[\alpha^j|X=x] - E[\alpha^j|X=x]| \right) \\ &\le M^k (2^k - 2)(n^{-1}\log n)^{1/2} C^{\sum_{j=1}^{k-1} j} \end{aligned} \tag{12}$$

Consider the event $A_{k,2}(x) = \{|\hat{E}[Y_t^k|X=x] - E[Y_t^k|X=x]| \le (n^{-1}\log n)^{1/2} C^{\sum_{j=1}^{k} j}\}$. If $C \ge 4M$ then for any k, $M^k(2^k-2) C^{\sum_{j=1}^{k-1} j} \le C^{\sum_{j=1}^{k} j}$ and also $C^{\sum_{j=1}^{k} j} > 4M^k$. Then it follows from Borel-Cantelli lemma and Bernstein's inequality in (11) that the event $A_{k,2}(x)$ occurs with probability 1 for all sufficiently large n. For this event we have

$$|\hat{\mu}_k(x_t) - \mu_k(x_t)| + |\hat{E}[\alpha^k|X = x] - E[\alpha^k|X = x]| \leq (n^{-1}\log n)^{1/2} C^{\sum_{l=1}^{k} l}$$

Finally, for any $\delta > 0$ we can choose $\eta = \frac{\delta}{\log C} > 0$ such that with probability 1,

$$\max_{1 \leq k \leq (\eta \log n)^{1/2}} \Big(|\hat{\mu}_k(x_t) - \mu_k(x_t)| + |\hat{E}[\alpha^k|X = x] - E[\alpha^k|X = x]| \Big) \leq n^{-1/2+\delta}$$

for all sufficiently large n.□

Proof of Theorem 3: It is easy to check that for any integer k, $\mu_k(x_t)$ is a homogeneous multivariate polynomial of degree k in $x_{t,1}, \ldots, x_{t,d}$. That is, it can be represented as

$$\mu_k(x_t) = \sum_{l_1 + \ldots + l_d = k} c_{l_1,\ldots,l_d}(k) x_{t,1}^{l_1} \ldots x_{t,d}^{l_d}. \tag{13}$$

Therefore, if for any k one can identify $\{c_{l_1,\ldots,l_d}(k), l_1 + \ldots + l_d = k\}$, then $\mu_k(x_t)$ is also identified for any x_t in the support of X_t. I will show that for any k, we can identify coefficients $c_{l_1,\ldots,l_d}(k)$ from a sequence of linear (in coefficients $c_{l_1,\ldots,l_d}(k)$) regressions.

Recall that $\mu_1(x_1) = x_1'\theta_\mu$ and $\mu_1(x_2) = x_2'\theta_\mu$, where θ_μ is identified since the matrix $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$ has full rank. Therefore, $\mu_1(x_1)$ and $\mu_1(x_2)$ are identified. Now suppose that for any $1 < j \leq k-1$ and any $x = (x_1, x_2)$ in the support of X, $\mu_j(x_1)$ and $\mu_j(x_2)$ are identified. Since U_1 and U_2 are independent conditional on X, then for any k,

$$E[(Y_2 - Y_1)^k|X = x] = \sum_{j=0}^{k} \binom{k}{j} (-1)^j \mu_{k-j}(x_1)\mu_j(x_2).$$

Therefore, we have:

$$\mu_k(x_1) + (-1)^k \mu_k(x_2) = E[(Y_2 - Y_1)^k|X = x] - \sum_{j=1}^{k-1} \binom{k}{j} (-1)^j \mu_{k-j}(x_1)\mu_j(x_2) \tag{14}$$

The right-hand side of equation (14) is identified since we assumed that $\mu_j(x_1)$ and $\mu_j(x_2)$ are identified for any $x \in \mathcal{X}$ and any $1 \leq j \leq k-1$. The right-hand side of equation (14) is linear in vector $\beta(k)$ whose typical element is $c_{l_1,\ldots,l_d}(k)$. That is, $\mu_k(x_1) + (-1)^k \mu_k(x_2) = z(k)'\beta(k)$, where $\dim(z(k)) = \binom{d+k-1}{d-1}$ and the typical element of the vector $z(k)$ is $x_{1,1}^{l_1} \ldots x_{1,d}^{l_d} + (-1)^k x_{2,1}^{l_1} \ldots x_{2,d}^{l_d}$. If X_t has at least $(d-1)$ continuous components, then the matrix $E[Z(k)Z(k)']$ has full rank for any fixed k, and therefore vector $\beta(k)$ is identified. In other words, $\{c_{l_1,\ldots,l_d}(k), l_1 + \ldots + l_d = k\}$ are identified, which together with expression in (13) implies that $\mu_k(x_1)$ and $\mu_k(x_2)$ are also identified for any $x \in \mathcal{X}$. $\Box$.

References

[1] 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(4), 379-397.

[2] Arellano, M., and S. Bonhomme (2009). "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," CEMMAP working paper CWP22/09.

[3] Beran, R., and P. Hall (1992). "Estimating Coefficient Distributions in Random Coefficient Regressions," *The Annals of Statistics*, 20(4), 1970-1984.

[4] Canay, I. (2010). "A Note on Quantile Regression for Panel Data Models," working paper, Northwestern University.

[5] Doksum, K. (1974). "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *The Annals of Statistics*, 2(2), 267-277.

[6] Galvao, A. (2008). "Quantile Regression for Dynamic Panel Data with Fixed Effects," working paper, University of Wisconsin-Milwaukee.

[7] Graham, B. S., J. Hahn, and J. L. Powell (2009). "A Quantile Correlated Random Coefficient Panel Data Model," working paper, University of California, Berkeley.

[8] Graham, B. S., and J. L. Powell (2008). "Identification and Estimation of 'Irregular' Correlated Random Coefficient Models," NBER Working Paper No. 14469.

[9] Greaves , (1982). “An Algorithm for the Hausdorff Moment Problem,” *Numerische Mathematik*, 39, 231-239.

[10] Koenker, R. (2004). “Quantile Regression for Longitudinal Data,” *Journal of Multivariate Analysis*, 91, 74-89.

[11] Lamarche, C. (2010). “Robust Penalized Quantile Regression Estimation for Panel Data,” *Journal of Econometrics*, forthcoming.

[12] Lasserre, J. B. (2002). “Bounds on Measures Satisfying Moment Conditions,” *Annals of Applied Probability*, 12(3), 1114-1137.

[13] Mnatsakanov, R. M., and A. S. Hakobyan (2009). “Recovery of Distributions via Moments,” in *Javier Rojo, ed., Optimality: The Third Erich L. Lehmann Symposium*, Beachwood, Ohio, USA: Institute of Mathematical Statistics, 252-265.

[14] Rosen, A. (2009). “Identification via Quantile Restrictions in Short Panels,” working paper, UCL.

[15] Ullah, A., and A. Pagan (1999). *Nonparametric econometrics*. Cambridge University Press.