Identification in Quantile Regression Panel Data Models with Fixed Effects and Small T

Maria Ponomareva*

Job Market Paper

University of Western Ontario

November 22, 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 periods each individual is observed is small. When fixed effects are pure location shifts, I show that the QR model is identified and suggest an estimator based on the recovering of a 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. Additionally, for a general class of random coefficients panel data models I show that those models are identified when all covariates are continuous, and show how one can identify variables with homogeneous individual responses.

1 Introduction

Quantile regression (QR) models are quite popular in the empirical literature: unlike traditional regression models that solely focus on the effect of covariates on the conditional mean of the outcome variable, quantile regression models allow to identify

*This is a preliminary version. 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.

and analyze other interesting features of the conditional distribution of the dependent variable. The attraction of quantile regression models is that they allow for possible heterogeneous effects of covariates: in many applications a researcher may have a reason to expect that the effect of covariates is not necessarily the same at different points of the distribution of outcome. For example, Abadie, Angrist and Imbens (2002) found that for women, Job Training Partnership Act (JTPA) training program had the largest effect at the low quantiles of earnings distribution. In yet another empirical study, Abrevaya and Dahl (2008) find that the effect of mother's smoking on child's birthweight is different for different quantiles of birthweight distribution.

If a researcher has panel data (when each member of the sample is observed in multiple point in time), she can control for some unobserved covariates that stay constant over time by utilizing panel data fixed effects models. The traditional linear panel data fixed effects panel data model for the conditional mean of the outcome variable works pretty straightforward even when there are only few observations per individual. The linear structure allows to difference out those unobserved covariates, and because the mean of the difference is the difference of the mean, looking at the mean relationship between change in outcome variable and change in covariates still allows a researcher to identify the effect of covariates on the conditional mean of outcome variable.

Considering the arguments outlined above, the combination of quantile regression models with panel data fixed effects models looks like a very attractive option. It allows us to kill two birds with one stone: study the causal effect of covariates on the distribution of outcome variable (quantile regression) while controlling for some unobserved heterogeneity (fixed effects). Combining individual heterogeneity (β_{it}) as a response of the outcome variable Y_{it} to a change in covariates X_{it} with individual fixed effects α_i's yields to the following random coefficient model:

$$Y_{it} = X_{it}'\beta_{it} + \alpha_i.$$

One more step is required to make this random coefficients model a linear quantile regression model: we assume that conditional quantiles of Y_{it} are linear in covariates. This assumption imposes a specific structure on random coefficients:

$$\beta_{it} = \theta(U_{it}),$$

where U_{it} is uniformly distributed on $[0,1]$ and the function $\tau \to x'\theta(\tau)$ is strictly increasing. This is the Doksum (1974) representation of the model with linear conditional quantiles.

In panel setups, this approach is indeed attractive and easy to implement when we care only about the mean effect of X_{it}, when we try to transfer it to a quantile regression setup, we run into this difficulty: we cannot difference out fixed effects any longer. The quantiles of the difference of two random variables do not equal to the difference of quantiles. That is, if we have a linear model and decide to difference out fixed effects, we end up with an object that (seemingly) tells us nothing about the conditional quantiles we are interested in. When the number of time observations for each individual in the sample is large, this is not a big issue: instead of differencing out fixed effects, we can treat them as parameters to be estimated. However, when each individual is observed only a only for a few times, this approach will not work.

In this paper I address the identification and estimation of a panel data linear quantile regression model with fixed effects when the number of observations for each individuals is small. In particular, I provide two sets of sufficient conditions that allow to identify QR panel data models with fixed effects under different assumption on the distribution of fixed effects. 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 these cases I present the condition under which the marginal quantile effects are identified when the number of time 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.

My main contribution in this paper is threefold. First, my identification result does not require the number of time periods (T) that we observe each individual in the sample to be large. Second, even if all covariates are discrete, my identification result does not require that we observe individuals for whom covariates do not change between time periods. Finally, the identification results in this paper are constructive and offer an estimation procedure that is based on estimating a sequence of moments.

The rest of this paper is organized as follows: Section 2 discusses the related literature. Section 3 presents the model and outlines identification and moments-based estimation strategy. Section 4 presents a set of assumptions sufficient for the identification of QR panel data model when the covariates can have discrete distribution and fixed effects are pure location shifts. Section 5 gives identification result for the case when the regressors have continuous distribution and also gives a set of sufficient conditions that allows to identify the QR model even when fixed effects are allowed to depend on quantile. Also, this section discusses possible extension of the proposed identification approach for identification of the distribution of random coefficients in a random coefficients panel data model. In particular, it suggests a test for homogeneous individual responses that is based on a $\sqrt{n}$-consistent estimators of first two moments. Section 6 explores some possible estimation procedures, and Section 7 concludes. All proofs of the results are collected in the Appendix.

2 Related Literature

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 (2011) 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 simply estimate fixed effects 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 (2010) 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. Evdokimov (2010) considers identification in a general class of nonparametric panel data models with unobservable heterogeneity that includes a linear quantile model with fixed effects. His identification and estimation result stems from the assumption that there are individuals in the sample for whom covariates do not change over time. However, this assumption may be too restrictive for some empirical applications. In particular, it does not allow to include year-specific effects.

A very interesting point of view on quantile regression panel data models is presented in Powell (2011). He changes the object of interest: instead of looking at the causal effect of covariates on the quantiles of the conditional distribution of an outcome (which is the object of interest here), he analyzes the quantiles of the *unconditional* distribution of an outcome and suggest a simple (and therefore attractive) moment-based approach to estimation of those unconditional quantiles.

In this paper I treat the QR panel data model as a special case of a random coefficients model. Heckman and Vytlacil (1998) emphasizes the importance of random coefficient models in capturing unobservable heterogeneity for some economic models. Beran and Hall (1992) and Beran, Feuerverger and Hall (1996) provide identification results for a cross-section random coefficients model. In particular, Beran and Hall (1992) show how one can identify and estimate the distribution of random coefficients if all the moments of this distribution are identified. Another interesting paper is by Fox et al. (2011), where the authors adopt a similar approach and show that the distribution of random coefficients in a logit model is identified by showing that all the moments of this distribution are identified. Finally, Hoderlein, Klemelä and Mammen (2007) propose a kernel based estimator for the joint probability density of the random coefficients that is based on the Radon transform.

Related papers that study random coefficients model in the context of panel data include Graham and Powell (2008) and Graham, Hahn and Powell (2009). The first paper looks at a certain features of the distribution of random coefficients, while the second paper looks at identification and estimation of certain conditional quantiles in a panel data model without fixed effects. 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.

3 The Model

If all conditional quantiles of the outcome variable Y_{it} are assumed to be linear in covariates X_{it}, the panel data quantile regression model can be represented as the following random coefficients model:

$$Y_{it} = X_{it}'\theta(U_{it}) + \alpha_i,\ i = 1, \ldots, n,\ t = 1, 2. \tag{1}$$

where $U_{it}|X_{i1}, X_{i2} \sim U[0,1]$ is individual- and time-specific error, and α_i is an additive individual effect that is the same for both time periods. The function $\tau \to x_t'\theta(\tau)$ is assumed to be strictly increasing on the interval $(0,1)$ for any given realization x_t in the support of X_{it}.[1]

A researcher observes (Y_{it}, X_{it}) in both time periods, but not U_{i1}, U_{i2} or α_i. If α_i's were observable, then conditional quantiles of Y_{it} would be

$$Q_{Y_{it}}(\tau|X_i = (x_1', x_2')', \alpha_i = \alpha) = x'\theta(\tau) + \alpha.$$

The parameter we are interested in is the vector function $\theta(\cdot)$. However, quantile functions are not linear functions and we cannot simply difference out fixed effects even when α_i's are independent from both X_i and U_{i1}, U_{i2}:

$$Q_{Y_{i2}-Y_{i1}}(\tau|X_i) \neq Q_{Y_{i2}}(\tau|X_i) - Q_{Y_{i1}}(\tau|X_i) \neq Q_{Y_{i2}}(\tau|X_i, \alpha_i) - Q_{Y_{i1}}(\tau|X_i, \alpha_i) = (X_{i1} - X_{i2})'\theta(\tau).$$

Our goal is to identify the so-called "quantile generating" function $\theta(\cdot)$ from the joint distribution of the observables: $(Y_{i1}, Y_{i2}, X_{i1}, X_{i2})$ while allowing for an arbitrary relationship between α_i and X_i (and possibly, between α_i and $U_i = (U_{i1}, U_{i2})$). If we treat model (1) as a random coefficient model, identification of $\theta(\cdot)$ amounts to identification of the distribution of the random coefficients. I show that under the certain assumptions this distribution is identified.

In particular, when fixed effects α_i's are independent from error terms U_i's, I show that both the conditional distribution of $X_{it}'\theta(U_{it})$ and the conditional distribution of α_i conditional on X_i are identified under certain conditions, and this identification argu-

[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.

ment is constructive. Also, I show that when all covariates are continuously distributed (which is a common assumption for identification of many random coefficients models), then we can identify $\theta(\cdot)$ even when fixed effects are not independent from error terms U_i's.

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_i. Once we obtain consistent estimators for this conditional distribution function, the inference procedure becomes really simple: we can sample from this distribution and estimate $\theta(\tau)$ for any given $0 < \tau < 1$ using the usual quantile regression technique (minimizing an appropriate function).

The standard error of such an estimator based on a sampling depends only on the standard error of the estimator 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. This approach is used e.g. in Beran and Hall (1992) to estimate distributions in a certain class of random coefficients regression models. I show that if either all covariates are discrete or all are continuous, these moments can be estimated at $\sqrt{n}$-rate. So, the estimation procedure for quantile coefficients can be summarized by the following two steps:

1. Estimate conditional moments $\left\{E\left((X_{it}'\theta(U_{it}))^k \,|X_i\right) : k = 1, 2, \ldots\right\}$ and recover the conditional distribution $F_t(u|X_i) = P(X_{it}'\theta(U_{it}) \leq u|X_i)$.

2. Sample from this conditional distribution (let's call this sample $(\tilde{Y}_{i1}, \tilde{Y}_{i2}, X_{i1}, X_{i2})$ and estimate $\theta(\cdot)$ by minimizing over θ the following objective function:

$$\sum_{t=1}^{2}\sum_{i=1}^{n}\rho_\tau\left(\tilde{Y}_{it} - X_{it}'\theta\right)$$

where $\rho_\tau(u) = u(\tau - 1(u < 0))$. Here $\tilde{Y}_{it}$'s are essentially equivalent to $(Y_{it} - \alpha_i)$'s.

As it was already mentioned above, the set of key identifying assumptions is different depending on whether the covariates have discrete or continuous support (in the latter case we can relax the assumption that fixed effects are independent from U_i's). 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.

4 Identification

In this section I present identification argument for the case when error terms U_i and fixed effect α_i are independent conditional on the observable covariates. Let $\mathcal{X} = suppX_i$ and let $x = (x_1', x_2')'$ denote a typical element in this set. Let's consider the following random variable:

$$Z_{it} = X_{it}'\theta(U_{it})$$

Identification of $\theta(\cdot)$ amounts to the identification of the conditional distribution of Z_{it} conditional on $X_i = (X_{i1}', X_{i2}')'$. If we can identify the conditional distribution of Z_{it} and if matrix $E(X_{it}X_{it}')$ has full rank, we can identify quantile coefficients $\theta(\tau)$ for $\tau \in (0, 1)$.

Assumption 1. *For any $x \in \mathcal{X}$, distributions of random variables $X_{it}'\theta(U_{it})|X_i = x$ for $t = 1, 2$ and distribution of $\alpha_i|X_i = x$ are uniquely determined by their moments (assuming that all moments exist and are 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, the conditional distributions of Z_{it} and α_i satisfy the Carleman's condition, so we can re-state Assumption 1 as

Assumption 2. *For any x in the support of X_i, the moments of Z_{it} conditional on $X_i = x$, denoted by $m_k(x, t) = E(Z_{it}^k|X_i = x)$ and the moments of α_i conditional on $X_i = x$, denoted by $a_k(x) = E(\alpha_i^k|X_i = x)$:*

(i) exist;

(ii) are finite;

(iii) satisfy the Carleman's condition for all $x \in \mathcal{X}$:

$$\sum_{k=1}^{\infty} (m_{2k}(x, t))^{-\frac{1}{2k}} = +\infty \text{ and } \sum_{k=1}^{\infty} (a_{2k}(x))^{-\frac{1}{2k}} = +\infty$$

Any distribution that satisfies the Carleman's condition is uniquely determined by its moments. The Carleman's condition is the *sufficient* condition for the determinacy of the Hamburger moment problem for distributions with unrestricted supports. If it is known that Z_{it} and α_i have compact supports, then the determinacy of the moment problem follows immediately without the need for the Carleman's condition.[2] The assumption of compact support for $Z_{it} = X_{it}'\theta(U_{it})$ is not that restrictive: the quantile regression literature often imposes the assumption that $\theta(\tau) \in \Theta$ for all $\tau \in \mathcal{T} \subset (0,1)$ where Θ and $\mathcal{T}$ are compact sets. If we strengthen this to $\theta(\tau) \in \Theta$ for all $\tau[0,1]$, we immediately get compact support of Z_{it} conditional on X_i.

Next assumption imposes some restrictions on the relationship between the observables X_i and the unobservables U_i and α_i in the model:

Assumption 3. *Random variables U_{i1}, U_{i2} and α_i are*

(i) independent conditional on X_i;

(ii) $U_{it} \sim U[0,1]$ for $t = 1, 2$.

Assumption 3 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 basically a location shift of the distribution of Z_{it}. Note that no conditions are imposed on the distribution of the vector of covariates $X_i = (X_{i1}', X_{i2}')'$: components of X_{it} can have a discrete or a continuous distribution, but X_{it} cannot include a constant term: $X_{it} \neq (1, \tilde{X}_{it}')'$.

I will use the following notation for the conditional distribution of Z_{it} and α_i:

$$m_k(x,t) = E(Z_{it}^k | X_i = x) \tag{2}$$

$$a_k(x) = E(\alpha_i^k | X_i = x) \tag{3}$$

The result below states that those moments are identified under Assumption 3.[3] The proof of this result is constructive and immediately suggest the way to estimate these moments.

Theorem 4.1. *Suppose that Assumption 3 is satisfied and assume 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 $x \in \mathcal{X}$, $m_k(x,1)$, $m_k(x,2)$ and $a_k(x)$ are identified.*

[2]With compact supports we have the Hausdorff moment problem. The determinacy of the Hausdorff moment problem follows from the Stone-Weierstrass theorem.

[3]All proofs, including this result, are collected in the Appendix.

Remark 4.1. The result in Theorem 4.1 can be extended to a general case of $T \geq 2$ periods, with the straightforward modification of Assumption 3 that allows to take into account more than just two time periods.

Remark 4.2. If we are not interested in the conditional distribution of α_i conditional on X_i, we do not have to require in Assumption 1 that all moments of this distribution exist: for the identification of $m_k(x,t)$ it is sufficient that only the first conditional moment of α_i exists and is finite.

Theorem 4.1 together with Assumption 1 imply the following corollary:

Corollary 4.2. *Suppose that conditions of Theorem 4.1 hold and that Assumption 1 is satisfied. Then for any x in the support of X_i, conditional distributions of $X_{it}'\theta(U_{it})$ and α_i conditional on $X_i = x$ are identified.*

For a traditional quantile regression model as introduced by Koenker and Bassett (1978) we know that the model is identified when the conditional distribution of the outcome variable Y conditional on observable covariates X is continuous. Otherwise, some quantiles may not be identified. In our case, the continuity of the conditional distribution of $Z_{it} = X_{it}'\theta(U_{it})$ (the "outcome" variable in which conditional distribution we are actually interested in) will be continuous if the vector function $\theta(\cdot)$ is continuous. This is summarized below by

Assumption 4. *The vector function of quantile slope coefficients $\theta(\cdot)$ is continuous everywhere on* $[0,1]$.

Note that the continuity of $\theta(\cdot)$ on the whole segment $[0,1]$ automatically gives us the compact support of the conditional distribution of Z_{it}, which in turn ensures that Assumption 1 is satisfied, as was noted above. Now we can state the main identification result:

Theorem 4.3. *Suppose that Assumptions 3 and 4 are satisfied. Also, assume that matrices $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$, $E(X_{i1}X_{i1}')$ and $E(X_{i2}X_{i2}')$ have full rank. Then the quantile regression model (1) is identified.*

The result in Theorem 4.3 tells us that the quantile panel data model (1) is identified under some restrictions on relationships between the unobservables U_{i1}, U_{i2}, α_i and the observables X_{i1} and X_{i2}. It works for both discrete and continuous covariates in X_{it}, and it does not rely on observing individuals for whom the covariates X_{it} do not change or almost do not change from period 1 to period 2.

One of the assumptions that is essential for the identification in Theorem 4.3 is the assumption that individual fixed effects α_i's are independent from U_{i1} and U_{i2} conditional on X_i. This may be a strong restriction on the assumed data generating process, and the next section shows that when all covariates are continuously distributed, the assumption of conditional independence of α_i and (U_{i1}, U_{i2}) can be removed.

5 Identification with Continuous Covariates

In this section I consider the same quantile regression panel data model (1), but now I will not impose the independence between α_i and U_{i1}, U_{i2}. Instead, I will put some restrictions on the distribution of X_{it}. Namely, if $d = \dim(X_{it})$, I require that at least $(d-1)$ covariates are continuously distributed, and that X_{it} does not include a constant term. The continuity of the distribution of covariates is one of the key identifying assumptions for the identification of certain random coefficients models (see, for example, Beran, Feuerverger and Hall (1996) or Fox et al. (2011a)).

Similar to the previous section, for any point $x \in \mathcal{X} = suppX_i$ I define a sequence of conditional moments of $Z_{it} = X_{it}'\theta(U_{it})$ as

$$m_k(x,t) = E[(X_{it}'\theta(U_{it}))^k \,|X_i = (x_1, x_2)].$$

Note that $m_k(x,t)$ is a 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 the random variable Z_{it}

$$m_k(x,t) = \sum_{l_1+\ldots+l_d=k} c_{l_1,\ldots,l_d}(k) x_{t,1}^{l_1} \cdot \ldots \cdot x_{t,d}^{l_d}, \tag{4}$$

where $x_t = (x_{t,1}, \ldots, x_{t,d})$. The fact that all conditional moments of Z_{it} have this very specific form help identify $m_k(x,t)$ from the sequence of conditional moments of the difference in outcomes between two time periods: $Y_{i2} - Y_{i1}$.

Assumption 3 can now be replaced with

Assumption 5. *Random variables U_{i1} and U_{i2} are* *(i) independent conditional on X_i;*

(ii) $U_{it} \sim U[0,1]$ for $t = 1, 2$.

Additionally, at least $(d-1)$ components of X_{it} are continuously distributed.

Unlike Assumption 3 employed in the previous section, Assumption 5 does not require fixed effects α_i's to be independent of $U_i = (U_{i1}, U_{i2})$ conditional on X_i.

Theorem 5.1. *Suppose that Assumptions 5 and 4 are satisfied. Also, assume that matrices $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$, $E(X_{i1}X'_{i1})$ and $E(X_{i2}X'_{i2})$ have full rank. Then the quantile regression model (1) is identified.*

The proof of Theorem 5.1 is constructive and suggests that one can estimate a sequence of moments $\{m_k(x_t), k \geq 1\}$ from a sequence of 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.

5.1 Random Coefficients Panel Data Model

A quantile regression model like (1) can be treated as a special case of a more general random coefficients (RC) panel data model with fixed effects:

$$Y_{it} = X'_{it}\beta_{it} + \alpha_i, \; i = 1, \ldots, n, \; t = 1, 2. \tag{5}$$

Here random coefficients β_{it} represent individual heterogeneity in the response of the outcome variable Y_{it} to changes in covariates X_{it}. Random coefficients models are very popular in empirical Industrial Organization literature, because such models allow for heterogeneity in marginal effects across individual agents. Beran, Feuerverger and Hall (1996) show identification of the distribution of random coefficients β_{it} for a linear cross-section RC model, and Fox et al. (2011) prove the identification of the RC Logit model. In both cases, the key identifying assumption is that all covariates are continuously distributed.

The identification result of Theorem 5.1 can be extended to the identification of the RC panel data model (6).

Assumption 6. *The distribution of β_{it} is uniquely determined by its moments (assuming that all moments exist and are finite).*

This assumption is similar to Assumption 1 and ensures that once all the moments of the distribution of β_{it} are identified, the distribution itself is also identified.

Assumption 7. *(i) Vectors of random coefficients β_{i1} and β_{i2} are independent conditional on X_i;*

(ii) Vector of covariates X_{it} is continuously distributed for $t = 1, 2$.

Assumption 7.(i) is somewhat restrictive as it does not allow for any relationship between individual responses in both periods: all intertemporal relationship between the outcomes in two periods is assumed to be captured by α_i. Assumption 7.(ii) is a standard identifying assumption in RC models.

Theorem 5.2. *If Assumptions 6 and 7 are satisfied by the model (1), and if matrix $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$ has full rank, then the distribution of random coefficients β_{it} is identified.*

5.1.1 Identification of Heterogeneous Effects: Dimension Reduction

Estimation of the joint distribution of random coefficients does not escape the curse of dimensionality that is common to all nonparametric estimators. One way to alleviate it is to assume that that individual responses to some of the covariates are homogeneous. That is, the model 6) can be written as

$$Y_{it} = X'_{1,it}\beta_{1,it} + X'_{2,it}\beta_2, \; i = 1, \ldots, n, \; t = 1, 2. \tag{6}$$

That is, individual responses to covariates in $X_{1,it}$ are heterogeneous ($\beta_{1,it}$ varies across individuals), but individual responses to covariates in $X_{2,it}$ are homogeneous (β_2 does not vary across individuals). It turns out that using the same moment-based approach we can identify the covariates with homogeneous individual responses and consequently test the null hypothesis that individual responses to some component of X_{it} are homogeneous against the alternative that they are heterogeneous. The good news is that this procedure does not suffer from the curse of dimensionality.

Theorem 5.3. *Suppose that Assumptions 6 and 7 are satisfied by the model (1). Then we need only first two moments to identify those components of β_{it} that are constant. Additionally, those moments can be estimated $\sqrt{n}$-consistently.*

This result immediately suggests a test procedure based on a $\sqrt{n}$-consistent estimator of the second central moment.

Remark 5.1. The proof of Theorem 5.3 does not require that all covariates are continuously distributed: it is sufficient that only the variable for which we want to test the homogeneity of individual responses is continuously distributed. Also, if some or all of the covariates are discrete but have rich enough support that allows identification of first two moments, we are able to identify the variables with homogeneous individual responses.

6 Estimation

As the identification results in Theorems 4.3 and 5.1 suggest, the problem of estimation of the panel data quantile regression model (1) can be separated into two steps:

1. Estimation of the conditional distribution of $X_{it}'\theta(U_{it})$;
2. Estimation of the quantile slope coefficients $\theta(\tau)$.

6.1 Estimation of the conditional distribution of $X_{it}'\theta(U_{it})$

The identification results of Theorems 4.3 and 5.1 are constructive: that is, they suggest a sequential procedure for estimation of moments $\{m_k(x,t), k = 1, \ldots\}$ for any x in the support of X_i, and then we can recover the distribution of a scalar random variable $x_t'\theta(U_{it})$ from its moments (the inverse Hamburger problem). The statistic literature offers a variety of techniques to recover the distribution from its moments. Beran and Hall (1992) provide a review of some of those methods, including approximations based on series and discrete approximations. When we assume conditional independence between fixed effect α_i and $U_i = (U_{i1}, U_{i2})$ conditional on X_i (Assumption 3), for a given sample size n it is possible to estimate (in theory) as many moments as we wish (see the proof of Theorem 4.1 in the Appendix). However, if we do not want to restrict possible relationship between α_i and U_i (Assumption 5) and when covariates are continuous, it is possible to estimate only a finite number of moments given the sample size n. In the following discussion I consider those two cases separately.

6.1.1 Estimation with Discrete Covariates When α_i and U_i Are Independent

Suppose that we observe the following data: $\{(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}), i = 1, \ldots, n\}$. When all covariates are discrete, the for any $k \geq 1$, we can $\sqrt{n}$-consistently estimate both $m_k(x,t)$ and $a_1(x)$ for any x in the support of X_i as the solution to[4]

$$\hat{m}_k(x,1) = \hat{E}_n[Y_1(Y_1 - Y_2)^{k-1}|x] - \sum_{j=1}^{k-1} \binom{k-1}{j} (-1)^j \hat{m}_{k-j}(x,1)\hat{m}_j(x,2) \tag{7}$$

$$- \hat{a}_1(x) \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j \hat{m}_{k-1-j}(x,1)\hat{m}_j(x,2) \tag{8}$$

and similarly,

$$\hat{m}_k(x,2) = \hat{E}_n[Y_2(Y_2 - Y_1)^{k-1}|x] - \sum_{j=1}^{k-1} \binom{k-1}{j} (-1)^j \hat{m}_{k-j}(x,2)\hat{m}_j(x,1) \tag{9}$$

$$- \hat{a}_1(x) \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j \hat{m}_{k-1-j}(x,2)\hat{m}_j(x,1), \tag{10}$$

where

$$\hat{m}_1(x,t) = x_t'\hat{\theta}_\mu, \tag{11}$$

$$\hat{a}_1(x) = \frac{1}{2}\left(\hat{E}_n[Y_1|X = x] + \hat{E}_n[Y_2|X = x] - \hat{m}_1(x,1) - \hat{m}_1(x,2)\right), \tag{12}$$

$$\hat{\theta}_\mu = \underset{\theta}{\operatorname{argmin}}\, \hat{E}_n(Y_1 - Y_2 - (X_1 - X_2)'\theta)^2. \tag{13}$$

For each fixed k, the estimators $\hat{m}_k(x,t)$ and $\hat{a}_1(x)$ defined above 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:

[4] Here I use the following notation: for a random sample $\{W_{it}, i = 1, \ldots, n, t = 1, 2\}$, $\hat{E}_n W_t = \frac{1}{n}\sum_{i=1}^n E_{it}$.

Assumption 8. *(i) The data* $(Y_{i1}, Yi2, X_{i1}, X_{i2})$ *are i.i.d. random sample from* $(\Omega, \mathcal{F}, P)$ *defined in (1).*

(ii) The support of $X_i = (X'_{i1}, X'_{i2})'$ *is discrete.*

Assumption 8.(i) is a standard random sampling condition, and Assumption 8.(ii) is the discrete support assumption that allows to estimate all moments at $\sqrt{n}$-rate.

Although we can estimate $m_k(x,t)$ for any k using the system of equations in (7), (9) and (11), the estimation error will accumulate as we increase k, and the standard deviation of the estimator of $m_k(x,t)$ for large k is going to be too high. One way to deal with this issue is to estimate only a finite number of moments and then invert this truncated sequence to get an estimator of the distribution of $X'_{it}\theta(U_{it})$ (the *truncated* Hamburger problem). The larger is the sample size n, the higher can be the number of moments that we can estimate with a reasonable error. Theorem 6.1 below specifies a rate condition for the number of moments to be estimated, $k(n)$ so that $\{\hat{m}_l(x,t), 1 \leq l \leq k(n)\}$ converge uniformly to $\{m_l(x,t).1 \leq l \leq k(n)\}$.

Theorem 6.1. *Suppose that conditions of Theorem 4.3 and Assumption 8 are satisfied. Then for any* $\delta > 0$ *there exists* $\eta > 0$ *such that with probability* 1*,*

$$\max_{1 \leq k(n) \leq (\eta \log n)^{1/2}} (|\hat{m}_k(x,t) - m_k(x,t)|) = O(n^{-1/2+\delta}) \text{ as } n \to \infty.$$

Remark 6.1. When some of the components of X_{it} are continuously distributed, we cannot estimate $\hat{m}_k(x,t)$ and $a_1(x)$ $n^{1/2}$-consistently any longer without imposing some strong assumptions. 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 6.1 will be adjusted accordingly to reflect the nonparametric convergence rate.

Theorem 6.1 allows us to estimate the distributions of $x'_t\theta(U_{it})$ for any given $x \in \mathcal{X}$ from its first $k(n)$ moments. Foe example, Beran and Hall (1992), Greaves (1982), and Mnatsakanov and Hakobyan (2009) provide several methods of solving the truncated Hamburger problem. 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 6.1 in the Appendix) suggests that one should choose k much smaller than the sample size n.

6.1.2 Estimation with Continuous Covariates When α_i and U_i May Not Be Independent

When at most only one regressor is discrete, Theorem 5.1 implies that the distribution of $\theta(U_{it})$ can be identified from the distribution of

$$Y_{i1} - Y_{i2} = X_{i1}'\theta(U_{i1}) - X_{i2}'\theta(U_{i2})$$

Here one can use either the discrete approximation in Fox and Kim (2011), or the Radon transform estimator proposed in Hoderlein, Klemelä and Mammen (2008). In particular, here I consider the discrete approximation in Fox and Kim (2011) and show how it reduces to the constrained Nonlinear Least Squares estimator in this case.

We want to approximate the conditional distribution of $X_{it}'\theta(U_{it})$ given X_i. Let $\theta(u) \in \Theta$, where Θ is a known compact parameter space and let $R(n)$ be the number of grid points for the discrete approximation. That is, we consider a grid $\Theta_{R(n)} \subset \Theta$ such that $\Theta_{R(n)} = \{\theta^1, \theta^2, \ldots, \theta^{R(n)}\}$. The grid is chosen by the researcher. Given the choice of the grid, the conditiona distribution of $X_{it}'\theta(U_{it})$ is approximated by the following discrete distribution:

$$X_{it}'\theta(U_{it}) = \begin{cases} X_{it}'\theta^1 & \text{with probability } p_1 \\ X_{it}'\theta^2 & \text{with probability } p_2 \\ \cdots & \cdots \\ X_{it}'\theta^{R(n)} & \text{with probability } p_{R(n)} \end{cases} \tag{14}$$

where $p_r \geq 0$ and $\sum_{r=1}^{R(n)} p_r = 1$. Our goal now is to estimate $p = (p_1, p_2, \ldots, p_{R(n)})'$.

Note that

$$\exp\{Y_{i1} - Y_{i2}\} = \exp\{X_{i1}'\theta(U_{i1})\} \cdot \exp\{-X_{i2}'\theta(U_{i2})\}$$

and

$$E\left(\exp\{Y_{i1} - Y_{i2}\} | X_i\right) = \left(\sum_{r=1}^{R(n)} p_r e^{X_{i1}'\theta^r}\right)\left(\sum_{r=1}^{R(n)} p_r e^{-X_{i2}'\theta^r}\right).$$

Then we can estimate vector p as:

$$\hat{p} = \underset{p}{\operatorname{argmin}} \left(\exp(Y_{i1} - Y_{i2}) - \left(\sum_{r=1}^{R(n)} p_r e^{X_{i1}'\theta^r} \right) \left(\sum_{r=1}^{R(n)} p_r e^{-X_{i2}'\theta^r} \right) \right)^2$$

$$\text{subject to } p_r \geq 0 \text{ and } \sum_{r=1}^{R(n)} p_r = 1 \ .$$

This is the nonlinear least squares problem, and a rate condition for $R(n)$ that insures uniform convergence is specified in Fox and Kim (2011) to satisfy

$$\frac{R(n) \log R(n)}{n} \to 0.$$

6.2 Estimation of the quantile slope coefficients $\theta(\tau)$

Suppose now that we picked some consistent estimator of the conditional distribution of $Z_{it} = X_{it}'\theta(U_{it})$ conditional on X_i. That is, let $\hat{F}_t(u|X_i = x)$ be a uniformly consistent estimator of $F_t(u|X_i = x) = P(Z_{it} \leq u|X_i = x)$. Our object of interest is

$$Q_{Z_{it}}(\tau|X_i) = X_i'\theta(\tau)$$

There is a simple procedure that allows us to estimate $\theta(\tau)$ for any $\tau \in (0,1)$: by drawing random samples from the distribution of Z_{it} and then applying standard quantile regression method to the new "data" (Z_{it}, X_{it}). This procedure can be summarized by these two steps:

1. Let $\{(Z_{it}, X_{it}) : i = 1, \ldots, m \text{ and } t = 1, 2\}$ be independent draws from the joint distribution of $(Z_{i1}, Z_{i2}, X_{i1}, X_{i2})$.

2. Then we can estimate $\theta(\tau)$ by

$$\hat{\theta}(\tau) = \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{m} \sum_{t=1}^{2} \rho_\tau \left(Z_{it} - X_{it}'\theta \right) \tag{15}$$

   where $\rho_\tau(u) = u(\tau - 1\{u \leq 0\})$ and where Θ is a compact parameter space[5] such that $\theta(\tau) \in \Theta$ for any $\tau \in [0, 1]$.

[5]The existence of such compact set Θ is implied by Assumption 4.

As long as $m \to \infty$ as $n \to \infty$ and $\hat{F}_t(\cdot|X_i = x)$ is a consistent estimator of $F(\cdot|X_i = x)$ for any x in the support of X_i, $\hat{\theta}(\tau)$ is a consistent estimator of $\theta(\tau)$ for any $\tau \in (0, 1)$. The convergence rate of $\hat{\theta}(\tau)$ is determined by the rate of convergence of $\hat{F}_t(\cdot|X_i = x)$. In particular, if $\hat{F}_t(\cdot|X_i = x)$ is $\sqrt{n}$-consistent, then $\hat{\theta}(\tau)$ is also $\sqrt{n}$-consistent.

7 Conclusion

This paper offers a novel approach to the identification and estimation of the linear quantile regression panel data models with fixed effects when the number of times T each individual is observed is small. This approach is based on the identification and estimation of moments of the conditional distribution of Y_{it} conditional on X_{it}.

In particular, I show that if fixed effects α_i are independent of individual time-specific shocks U_{it}, the model is fully identified including the conditional distribution of fixed effects. Also, when all covariates are discrete, the moments of those conditional distributions can be estimated at a parametric rate. When the covariates are continuously distributed, I show that the quantiles of conditional distribution of Y_{it} can be identified even when fixed effects are allowed to depend on individual time-specific shocks U_{i1} and U_{i2}.

The identification result in the paper can also be applied to a more general class of random coefficients panel data models. I also show how one can use the approach developed in this paper to test homogeneity of individual responses in such models.

Estimation of conditional quantiles $\theta(\cdot)$ here relies on the estimation of the conditional distribution of $X_{it}'\theta(U_{it})$ conditional on X_i: once those conditional distributions can be uniformly estimated, the estimation of the quantile slope coefficients is relatively simple. To estimate those conditional distributions, I suggest some procedures based on a discrete approximation result in Fox and Kim (2011). However, such an approach may be computationally intensive if the number of explanatory variables in the model is high. A more extensive treatment of inference in such models is left for future research.

A Appendix

A.1 Proof of Theorem 4.1

For the ease of presentation, index i is omitted here. Let $\mathcal{X} = suppX_i$ and consider any $x = (x_1, x_2) \in \mathcal{X}$. I'll be using the following notation: for any random variable W,

$$E(W|x) = E(W|X = x).$$

By Assumption 3, U_t and α are independent conditional on $X = x$. Note that $m_1(x, t) = x_t\theta_\mu$, where $\theta_\mu = E(\theta(U_{it})$. Therefore, $m_1(x, t)$ is identified if θ_μ is identified. But the latter is identified from

$$E[Y_2 - Y_1|X = (x_1, x_2)] = (x_2 - x_1)'\theta_\mu,$$

if matrix $E[(X_2 - X_1)(X_2 - X_1)']$ has full rank.
This implies that $a_1(x)$ is also identified from

$$a_1(x) = E[\alpha|X = x] = \frac{1}{2}\left(E[Y_1|X = x] + E[Y_2|X = x] - m_1(x, 1) - m_1(x, 2)\right).$$

Suppose now that for any $j \in \mathbb{N}$, $1 \le j \le k-1$ both $a_j(x)$ and $m_j(x, t)$ are identified and let's define $a_0(x) = 1$ and $m_0(x, t) = 1$. I will show that then $a_k(x)$ and $m_k(x, t)$ are also identified.
In order to do this, consider

$$\begin{aligned} E[Y_1(Y_1 - Y_2)^{k-1}|x] &= E[Y_1(x_1\theta(U_1) - x_2\theta(U_2))^{k-1}|x] \\ &= E\left[(x_1\theta(U_1) + \alpha)\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\left(x_1'\theta(U_1)\right)^{k-1-j}\left(x_2'\theta(U_2)\right)^j|x\right] \\ &= m_k(x, 1) + \sum_{j=1}^{k-1}\binom{k-1}{j}(-1)^j m_{k-j}(x, 1)m_j(x, 2) \\ &\quad + a_1(x)\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j m_{k-1-j}(x, 1)m_j(x, 2) \end{aligned} \tag{16}$$

where the last equality follows from independence of U_1 and U_2 and α_i conditional on x (Assumption 3.(ii)).
Then

$$m_k(x,1) = E[Y_1(Y_1 - Y_2)^{k-1}|x] - \sum_{j=1}^{k-1} \binom{k-1}{j} (-1)^j m_{k-j}(x,1) m_j(x,2) \\ - a_1(x) \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j m_{k-1-j}(x,1) m_j(x,2)$$

and similarly,

$$m_k(x,2) = E[Y_2(Y_2 - Y_1)^{k-1}|x] - \sum_{j=1}^{k-1} \binom{k-1}{j} (-1)^j m_{k-j}(x,2) m_j(x,1) \\ - a_1(x) \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j m_{k-1-j}(x,2) m_j(x,1)$$

are identified if all $m_j(x,t)$ are identified up to $j \leq k-1$ and if $a_1(x)$ is identified.

Since the choice of k is arbitrary, we showed that all the moments are identified for the conditional distributions of Z_{it}.

Additionally, we can identify all the moments of the conditional distribution of α_i as

$$a_k(x) = E[Y_1^{k-1} Y_2 | X = x] - m_1(x,2) \sum_{j=0}^{k-1} \binom{k-1}{j} m_{k-1-j}(x,1) a_j(x) \\ - \sum_{j=0}^{k-2} \binom{k-1}{j} m_{k-1-j}(x,1) a_{j+1}(x).$$

$\square$

A.2 Proof of Theorem 4.3

Assumption 3 together with the assumption that matrix $E[(X_{i2} - X_{i1})(X_{i2} - X_{i1})']$ has full rank imply that all moments of the conditional distribution of $Z_{it} = X_{it}'\theta(U_{it})$ are identified (Theorem 4.1).

Continuity Assumption 4 ensures that Assumption 1 is satisfied, which together with identification of all conditional moments of Z_{it} ensures that the conditional distribution of Z_{it} is identified (Corollary 4.2).

Finally, when matrix $E(X_{it}X_{it}')$ has full rank, for any $\tau \in [0,1]$, $\theta(\tau)$ is a unique solution to the following minimization problem:

$$\theta(\tau) = \underset{\theta}{\operatorname{argmin}}\, E\rho_\tau(Z_{it} - X_{it}'\theta) \tag{17}$$

where $\rho_\tau(u) = u(\tau - 1(u < 0))$. Since the conditional distribution of Z_{it} given X_{it} is identified, the objective function in (17) is also identified. This, in turn, implies that the solution to (17) is identified. That is, for any $\tau \in [0,1]$, $\theta(\tau)$ is identified. $\square$

A.3 Proof of Theorem 5.1

We only need to show that all moments $m_k(x,t)$ are identified under the conditions of Theorem 5.1. Then the proof of identification of quantile coefficients $\theta(\cdot)$ is exactly the same as in Theorem 4.3. Again, index i is omitted here.

For any k, $m_k(x,t)$ is a homogeneous polynomial of degree k in $x_{t,1},\ldots,x_{t,d}$. That is, it can be represented as

$$m_k(x,t) = \sum_{l_1+\ldots+l_d=k} c_{l_1,\ldots,l_d}(k) x_{t,1}^{l_1}\ldots x_{t,d}^{l_d}. \tag{18}$$

That is, it is sufficient to show that under the conditions of Theorem 5.1 the coefficients $\{c_{l_1,\ldots,l_d}(k), l_1+\ldots+l_d = k\}$ are identified for any k. I will show now that we can identify those coefficients from a sequence of linear (in coefficients $c_{l_1,\ldots,l_d}(k)$) regressions.

When matrix $E[(X_{i2}-X_{i1})(X_{i2}-X_{i1})']$ has full rank, we can identify $m_1(x,t)$ (see the proof of Theorem 4.3). Now suppose that for any $1 < j \leq k-1$ and any $x = (x_1, x_2)$ in the support of X, $m_j(x,1)$ and $m_j(x,2)$ are identified. Since U_1 and U_2 are independent conditional on X, then for any k

$$E[(Y_2 - Y_1)^k | X = x] = \sum_{j=0}^{k} \binom{k}{j} (-1)^j m_{k-j}(x,1) m_j(x,2).$$

Therefore, we have:

$$m_k(x,1) + (-1)^k m_k(x,2) = E[(Y_2 - Y_1)^k | X = x] - \sum_{j=1}^{k-1} \binom{k}{j} (-1)^j m_{k-j}(x,1) m_j(x,2) \tag{19}$$

The right-hand side of equation (19) is identified since we know $m_j(x_1)$ and $m_j(x_2)$ for any $x \in \mathcal{X}$ and any $1 \le j \le k-1$. The right-hand side of equation (19) is linear in vector $C(k)$ whose typical element is $c_{l_1,\ldots,l_d}(k)$. That is,

$$m_k(x,1) + (-1)^k m_k(x,2) = w(k)' C(k),$$

where $\dim(w(k)) = \binom{d+k-1}{d-1}$ and the typical element of the vector $w(k)$ is

$$x_{1,1}^{l_1} \ldots x_{1,d}^{l_d} + (-1)^k x_{2,1}^{l_1} \ldots x_{2,d}^{l_d}$$

When X_t has at least $(d-1)$ continuous components and there is no constant term in X_{it}, the matrix $E[W(k)W(k)']$ has full rank for any k, and therefore vector $C(k)$ is identified. In other words, $\{c_{l_1,\ldots,l_d}(k), l_1 + \ldots + l_d = k\}$ are all identified, which immediately implies that $m_k(x,1)$ and $m_k(x,2)$ are also identified for any $x \in \mathcal{X}$.
The rest of the proof follows the proof of Theorem 4.3. $\square$

A.4 Proof of Theorem 5.2

Arguments similar to those used in the proof of Theorem 5.1 can be used here to show that for each k, $E[(X_{it}'\beta_{it}|X_i]$ is identified. Together with Assumption 6 this implies that the conditional distribution of $'_{it}\beta_{it}$ conditional on X_i is identified. The identification result for the distribution of β_{it} immediately follows from Beran, Feuerverger and Hall (1996). $\square$

A.5 Proof of Theorem 5.3

Without loss of generality, assume that $X_{it} = (X'_{1,it}, X'_{2,it})'$, $\beta_{it} = (\beta'_{1,it}, \beta'_{2,it})'$ and that $\beta_{2,it} \equiv \beta_2$. But $\beta_{2,it} \equiv \beta_2$ if and only if for any X_i,

$$\begin{aligned} 0 &= Var[X'_{2,it}\beta_{2,it}|X_i] \\ &= E[(X'_{2,it}\beta_{2,it})^2|X_i] - \left(E[X'_{2,it}\beta_{2,it}|X_i]\right)^2 \end{aligned}$$

Here $E[X'_{2,it}\beta_{2,it}|X_i]$ can be easily identified and estimated $\sqrt{n}$-consistently from a linear regression of $Y_{i1} - Y_{i2}$ on $X_{i1} - X_{i2}$.

The proof of Theorem 5.1 suggests that the second moment $E[(X'_{it}\beta_{it})^2|X_i]$ is also identified and can be $\sqrt{n}$-consistently estimated from the appropriate linear regression. But

$$E[(X'_{it}\beta_{it})^2|X_i] = E[(X'_{1,it}\beta_{1,it})^2|X_i] + E[(X'_{2,it}\beta_{2,it})^2|X_i] + 2E[(X'_{1,it}\beta_{1,it})(X'_{2,it}\beta_{2,it})|X_i] \tag{20}$$

All three components of the right-hand side in (20) are polynomials of degree 2 in components of only $X_{1,it}$, only $X_{2,it}$ and both $X_{1,it}$ and $X_{2,it}$, correspondingly. That is, $E[(X'_{2,it}\beta_{2,it})^2|X_i]$ is identified and can be estimated $\sqrt{n}$-consistently. We can use these estimators$\square$

A.6 Proof of Theorem 6.1

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 8(ii), $\mathcal{X}$ is a finite set. Therefore, it is sufficient to show that the claim is true for a given $x \in \mathcal{X}$. The proof also includes uniform estimation of the conditional moments of α_i.

Recall that

$$\begin{aligned} \hat{a}_1(x) &= \tfrac{1}{2}\left(\hat{E}_n[Y_1|X=x] + \hat{E}_n[Y_2|X=x] - \hat{m}_1(x,1) - \hat{m}_1(x,2)\right), \\ \hat{m}_k(x,1) &= \hat{E}_n[Y_1(Y_1-Y_2)^{k-1}|x] - \sum_{j=1}^{k-1}\tbinom{k-1}{j}(-1)^j \hat{m}_{k-j}(x,1)\hat{m}_j(x,2) \\ &\quad -\hat{a}_1(x)\sum_{j=0}^{k-1}\tbinom{k-1}{j}(-1)^j \hat{m}_{k-1-j}(x,1)\hat{m}_j(x,2) \end{aligned}$$

Define the following set

$$A_{k,1}(x) = \left\{ |\hat{m}_j(x,t) - m_j(x,t)| + |\hat{a}_j(x) - a_j(x)| \leq (n^{-1}\log n)^{1/2} C^{\sum_{l=1}^{j} l},\ 1 \leq j \leq k-1 \right\} \tag{21}$$

We have $m_k(x,t) = E[Y_t^k | X = x] - \sum_{j=1}^{k} \binom{k}{j} m_{k-j}(x,t) a_j(x)$, so that

$$\hat{m}_k(x,t) + \hat{a}_k(x) = \hat{E}_n[Y_t^k | X = x] - \sum_{j=1}^{k-1} \binom{k}{j} \hat{m}_{k-j}(x,t)\hat{a}_j(x),$$

and therefore

$$\begin{aligned} &|\hat{m}_k(x,t) - m_k(x,t)| + |\hat{a}_k(x) - a_k(x)| \leq |\hat{E}_n[Y_t^k | X = x] - E[Y_t^k | X = x]| \\ &+ \sum_{j=1}^{k-1} \binom{k}{j} |\hat{m}_{k-j}(x,t)\hat{a}_j(x) - m_{k-j}(x,t) a_j(x)| \end{aligned} \tag{22}$$

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

$$P\{|\hat{E}_n[Y_t^k | X = x] - E[Y_t^k | X = x]| \geq n^{-1/2} M^k s\} \leq 2e^{-s^2/4} \tag{23}$$

Note that the second term of the right-hand side of (22) is bounded by:

$$\begin{aligned} &\sum_{j=1}^{k-1} \binom{k}{j} |\hat{m}_{k-j}(x,t)\hat{a}_j(x) - m_{k-j}(x_t) a_j(x)| \\ &\leq M^k \sum_{j=1}^{k-1} \binom{k}{j} \left(|\hat{m}_j(x,t) - m_j(x,t)| + |\hat{a}_j(x) - a_j(x)| \right) \\ &\leq M^k (2^k - 2)(n^{-1}\log n)^{1/2} C^{\sum_{j=1}^{k-1} j} \end{aligned} \tag{24}$$

Now let's consider the event $A_{k,2}(x) = \{|\hat{E}_n[Y_t^k | X = x] - E[Y_t^k | X = x]| \leq (n^{-1}\log n)^{1/2} C^{\sum_{j=1}^{k} j}\}$. If $C \geq 4M$ then for any k, $M^k(2^k - 2) C^{\sum_{j=1}^{k-1} j} \leq C^{\sum_{j=1}^{k} j}$ and also $C^{\sum_{j=1}^{k} j} > 4M^k$. Then it follows from Borel-Cantelli lemma and Bernstein's inequality in (23) that the event $A_{k,2}(x)$ occurs with probability 1 for all sufficiently large n. For this event we have

$$|\hat{m}_k(x,t) - m_k(x,t)| + |\hat{a}_k(x) - a_k(x)| \leq (n^{-1}\log n)^{1/2} C^{\sum_{l=1}^{k} l}$$

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

$$\max_{1 \leq k \leq (\eta \log n)^{1/2}} \left(|\hat{m}_k(x,t) - m_k(x,t)| + |\hat{a}_k(x) - a_k(x)|\right) \leq n^{-1/2+\delta}$$

for all sufficiently large n. □

References

[1] Abadie, A., J. Angrist, and G. Imbens (2002) "Instruments Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," Econometrica, 70(1), 91-117.

[2] 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.

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

[4] Beran, R., A. Feuerverger, and P. Hall. (1996). On Nonparametric Estimation of Intercept and Slope Distributions in Random Coefficient Regression," *Annals of Statistics*, 24 2569-2692.

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

[6] Canay, I. (2011). "A Simple Approach to Quantile Regression for Panel Data," *The Econometrics Journal*, forthcoming.

[7] 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.

[8] Evdokimov, K. (2010). "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," workin paper, Princeton University.

[9] Fox, J.T., K. Kim, S. Ryan, and P. Bajari (2011). "The Random Coefficients Logit Model Is Identified," *Journal of Econometrics*, forthcoming.

[10] Fox, J.T., and K. Kim (2011). "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models," working paper, University of Michigan.

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

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

[13] Graham, B. S., and J. L. Powell (2011). "Identification and Estimation of Average Partial Effects in 'Irregular' Correlated Random Coefficient Models," working paper, University of California, Berkeley.

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

[15] Heckman, J., and E. Vytlacil (1998). "Instrumental Variables Methods for the Correlated Random Coefficients Model: Estimating the Average Return to Schooling when the Return is Correlated with Schooling," *Journal of Human Resources*, 33, 974-987.

[16] Hoderlein, S., J. Klemelä, and E. Mammen (2008). "Analyzing the Random Coefficient Model Nonparametrically", working paper, Brown University.

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

[18] Koenker, R., and G. Bassett (1978). " Regression Quantiles," *Econometrica*, 46, 33-50.

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

[20] 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.

[21] Powell, D. (2011). "Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors," working paper, RAND

[22] Rosen, A. (2010). "Set Identification via Quantile Restrictions in Short Panels," working paper, UCL.

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