Identification in Some Random Coefficients Panel Data Models (With Application to Quantile Regression)

Maria Ponomareva*

Northern Illinois University

May 2014

Abstract

This paper considers a random coefficients panel data model with individual-specific intercepts (or fixed effects). The identification of the distribution of random slope coefficients is established in two settings: when random slope coefficients are conditionally independent from individual-specific intercepts; and when individual-specific intercepts are allowed to depend on random slope coefficients, too. The identification result requires only two observations per each unit. The paper considers a quantile regression panel data model with fixed effects as a special case of random coefficients panel data model and provides sufficient conditions under which conditional quantiles functions are identified. All identification results are constructive, and estimation procedure based on these results is proposed.

1 Introduction

Two of the likely most popular methods of controlling for the unobserved heterogeneity are (1) individual-specific intercept (or "fixed effects") that take into account omitted

*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.

variables that stay constant over time; and (2) allowing for heterogeneous responses across individuals and then studying the distribution of random responses in the population (random coefficients models). This paper combines both approaches and utilizes the availability of multiple observations per each individual to show that under certain restrictions on the joint distribution of random coefficients and individual-specific intercepts, the distribution of random coefficients can be identified and estimated using only two observations for each unit. The main model considered is the paper is as follows: for individual i in time period t, the outcome y_{it} for is generated as

$$y_{it} = x_{it}'\beta_{it} + \alpha_i \tag{1}$$

where $\beta_{it} \in \mathbb{R}^k$ is a vector of individual i's slope coefficients, and α_i is an additive individual-specific effect that stays the same in both time periods. For example, one can think of α_i as an individual-specific intercept that does not change with time. We assume that only a small number T of observations per individual is available, and without loss of generality we set $T = 2$. So, the researcher observes n i.i.d. random draws $(y_{i1}, y_{i2}, x_1, x_2)$, and the goal is to identify the distribution of random coefficients β_{i1} and β_{i2}.

A special case of the above model is a panel data quantile regression model:

$$y_{it} = x_{it}'\theta(u_{it}) + \alpha_i \tag{2}$$

where u_{it} is a scalar $U[0, 1]$ random variable that determines the ranking of individual i in the distribution, and mapping $\tau \mapsto x_{it}\theta(\tau)$ is a strictly increasing function for any x_{it} in the support. In this case, the conditional quantiles of y_{it} are given by

$$Q_\tau(y_{it}|x_{it}, \alpha_i) = \alpha_i + x_{it}'\theta(\tau)$$

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 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.

This paper addresses the identification and estimation of a linear random coefficient panel data model with fixed effects when the number of observations for each individuals is small. The identification strategy relies on the following key points:

- The ability to identify (and estimate consistently) mean effects of covariates x_{it}'s and the mean of the individual-specific effect, α_i. This can be done using a combination of a standard identification and estimation of a linear panel data model with fixed effects and identification of nonparametric conditional mean of y_{it} given $x_i = (x_{i1}, x_{i2})$ and hence the conditional mean of α_i given x_i.
- Either:
  - deconvolution of $x_{it}'\beta_{it}$ and α_i using the joint distribution of outcomes in both time periods;
  - or deconvolution of $x_{i1}'\beta_{i1}$ and $x_{i2}'\beta_{i2}$ from the distribution of the difference in the outcomes $\Delta y_i = y_{i2} - y_{i1}$. This approach relies on the assumption that one is able to shift the difference in covariates, $\Delta x_i = x_{i2} - x_{i1}$ freely relative to the first (or second) period outcome x_{i1} (x_{i2}).
- Identification of necessary characteristics of the distribution of β_{it} from conditional distributions of linear combinations $x_{it}'\beta_{it}$. For example, conditional quantiles are identified from conditional distribution of $x_{it}'\beta_{it}$; while first and second moments of β_{it} do not require identification of the conditional distribution. Rather, those moments are identified from a sequence of moment conditions related to Δy_i and can be estimated at a parametric rate.

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 the quantile regression panel data model is treated 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 (2012) and Graham, Hahn and Powell (2009). The first paper looks at a certain feature of the distribution of random coefficients (average partial effects), while the second paper looks at identification and estimation of 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 some 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 discusses identification assumptions and presents identification results for two main set ups: when β_{i1}, β_{i2} and α_i are mutually independent conditional on x_i and when only β_{i1} and β_{i2} are independent conditional on x_i (thus this case allows fixed effects to be correlated with random coefficients). Second part of this section discusses the conditions for identification of homogeneous effects (these conditions are weaker than the conditions for the identification of the distribution of $x'_{it}\beta_{it}$). Third part of Section 2 discusses identification of conditional quantiles of the distribution of $x'_{it}\beta_{it}$ under linear conditional quantiles assumption. Section 3 explores some possible estimation procedures, and Section 4 concludes. All proofs of the results are collected in the Appendix.

2 Identification

In what follows, subscript i is omitted (excluding estimation results). We begin by considering the set of assumptions that provides identification of the conditional distribution of $x_t'\beta_t$ without relying of continuity of the distribution of $\Delta x = x_2 - x_1$:

Assumptions ID.1 :

- **A1** Conditional independence: conditional on x, β_1, β_2 and α are mutually independent.
- **A2** Independence and stationary mean: β_t is independent from x and $E[\beta_t|x] = \beta_\mu$ for $t = 1, 2$.
- **A3** Full rank: matrix $E[(x_2 - x_1)(x_2 - x_1)']$ has full rank.
- **A4** Non-zero characteristic functions: for $t = 1, 2$ and all $x \in \mathcal{X}$ and $s \in (-\infty, +\infty)$, the characteristic functions $\varphi_{x_t'\beta_t|x}(s) \neq 0$ and $\varphi_{\alpha|x}(s) \neq 0$
- **A5** Continuous covariates: x_1 and x_2 are continuously distributed, and a unit ball $U_0 = \{x_t \in \mathbf{R}^k : \|x_t\| \leq 1\}$ is contained in the support of x_t, for $t = 1, 2$.
- **A6** Random sampling: $T = 2$ and $\{(y_{i1}, y_{i2}, x_i), i = 1, \ldots, n\}$ are n i.i.d. random draws from the DGP.

Here assumptions A1 and A2 are the key identifying assumptions: assumption A2 allows us to identify the mean of β_t from the linear panel data model with fixed effects. For example, A2 holds when β_1 and β_2 are independent from $x = (x_1, x_2)$ (standard uncorrelated random coefficients assumption) and identically distributed (stationarity). Assumption A1 is strong since it requires fixed effect α to be conditionally independent from random effects β_t's. However, it is possible to relax this assumption under additional support conditions. Assumption A3 is a standard full rank condition for the identification of the linear panel data model with fixed effects. Assumption A4 is a technical condition (see e.g. Evdokimov (2010)). Assumption A5 allows us to apply Cramér-Wold device to identify the joint distribution of random coefficients. This assumption is quite strong in the sense that it requires that basically any combination of positive and negative components of x_t has a positive density (the normalization to a unit ball is purely technical and not restrictive). However, without this assumption the identification of the distribution of β_t would be difficult. An alternative would be to impose some conditions on the shape of the conditional distribution of $x_t'\beta_t$ given x: for example, linear conditional quantile functions is one such restriction.

Theorem 2.1. *Suppose that assumptions A1 - A4 are satisfied. Then the conditional density of $x_t'\beta_t$, $f_{x_t'\beta_t|x}(\cdot|x)$, and the conditional density of α, $f_{\alpha|x}(u|x)$ are identified.*

The identification of the distribution of random coefficients, β_t, follows immediately from Theorem 2.1 via the application of CramérWold device.

Corollary 2.2. *Suppose that assumptions A1 - A4 and A5 hold. Then the distribution of β_t is identified for $t = 1, 2$.*

The mutual independence assumption A1 is a strong assumption. However, it can be relaxed somewhat at the cost of stronger support and stationarity assumptions. In particular, consider the following alternative (to Assumptions ID.1) set of assumptions:

Assumptions ID.2 :

- **A1′** Conditional independence: conditional on x, β_1 and β_2.
- **A2′** Stationarity: β_1 and β_2 are identically distributed. Also, $E(\beta_t|x) = \beta_\mu$ exists, and for any x_t, $E(x_t'\beta_t)^2 < \infty$.
- **A3′** For each x_1 there exists $h \neq -1$ such that $\Delta x = x_2 - x_1$ is continuously distributed in an open area $U_a(hx_1) = \{\Delta x \in \mathbf{R}^k : \|\Delta x - h^* x_1\| < a\}$ for some $a > 0$ with the density that is positive everywhere in $U_a(hx_1)$.
- **A4′** Non-zero characteristic functions: for $t = 1, 2$ and all $x \in \mathcal{X}$ and $s \in (-\infty, +\infty)$, the characteristic functions $\varphi_{x_t'\beta_t|x}(s) \neq 0$ and $\varphi_{\alpha|x}(s) \neq 0$
- **A5′** Continuous covariates: x_1 is continuously distributed, and a unit ball $U_0 = \{x_1 \in \mathbf{R}^k : \|x_1\| \leq 1\}$ is contained in the support of x_1.
- **A6′** Random sampling: $T = 2$ and $\{(y_{i1}, y_{i2}, x_i), i = 1, \ldots, n\}$ are n i.i.d. random draws from the DGP.
- **A7′** The characteristic function

  $$\varphi_{\Delta y|(x_1,(1+h)x_1)}(s) = E\left[\exp\left(\mathrm{i}s(y_2 - y_1)\right)|x = (x_1, (1+h)x_1)\right]$$

  is twice continuously differentiable in h everywhere in an open ball around $(x_1, (1+h)x_1)$.

Here assumption A1′ is weaker than assumption A1 since it does not require independence between β_t and α. However, assumption A2′ is stronger then A2. Assumption A3′ now requires the difference in covariates between two time periods to be continuously distributed. This condition rules out discrete change in covariates. However, it does not require Δx to have the unbounded support. Also, condition that $h \neq -1$ is not restrictive: it rules out the trivial case when $x_2 = -x_1$ and $f_{x_2,x_1}(x,-x) > 0$. Assumption A7$'$ is a technical assumption that allows consistent estimation of the derivative of $\varphi_{\Delta y|(x_1,(1+h)x_1)}(s)$ w.r.t. h. The rest of the assumptions remain the same as before.

The identification in this case relies on the difference between the outcome in two time periods. Assumption A3$'$ essentially allows us to shift the second period outcome freely having first period outcome fixed. The identification result is summarized in the theorem below.

Theorem 2.3. *Suppose that assumptions A1$'$ - A4$'$ are satisfied. Then the conditional density of $x_t'\beta_t$, $f_{x_t'\beta_t|x}(\cdot|x)$ for $t = 1, 2$ is identified.*

Again, the identification of the distribution of random coefficients, β_t, follows immediately from Theorem 2.3:

Corollary 2.4. *Suppose that assumptions A1$'$ - A4$'$ and A5$'$ hold. Then the distribution of β_t is identified for $t = 1, 2$.*

2.1 Identification of Homogeneous Effects

Estimation of the joint distribution of random coefficients does not escape the curse of dimensionality since we need to estimate the conditional characteristic function of Δy nonparametrically. One way to alleviate this problem is to assume that that individual responses to some of the covariates are homogeneous. In particular, suppose that the model 1) can be written as

$$y_{it} = (x_{it}^{(1)})'\beta_{1,it} + (x_{it}^{(2)})'\delta, \; t = 1, 2. \tag{3}$$

Here $x_{it} = ((x_{it}^{(1)})', (x_{it}^{(2)})')'$ and $\beta_{it} = (\beta_{1,it}', \delta')'$, That is, individual responses to covariates in $x_{it}^{(1)}$ are heterogeneous ($\beta_{1,it}$ varies across individuals), but individual responses to covariates in $x_{it}^{(2)}$ are homogeneous (δ does not vary across individuals).

The response to covariates in $x_{it}^{(2)}$ will be homogeneous when conditional on x_i, the variance (a second central moment) of $(x_{it}^{(2)})'\delta$ is zero. Fortunately, the identification of first and second moments of β_{it} does not require all covariates to be continuous, as summarized by the following result.

Theorem 2.5. *Suppose that assumptions 2 and 2 hold for the model (1). Also assume that the following rank conditions hold:*

(i) Let $w_{jm} = x_{j,1}x_{m,1} + x_{j,2}x_{m,2}$ *where* $j, m = 1, \ldots, k$ *and* $m \leq j$*, so that* w *is a* $1 \times k + k(k-1)/2$ *vector composed of* w_{jm}*. A* $k + (k(k-1)/2) \times (k + k(k-1)/2)$ *matrix* $E[ww']$ *has full rank.*

(ii) Let $v_t = (x_{1,t}^2, \ldots, x_{k,t}^2, x_{1,t}x_{2,t}, \ldots, x_{j,t}x_{m,t}, \ldots, x_{k,t}x_{k-1,t})'$ *be a* $k+k(k-1)/2$ *vector of cross-products of elements of* x_t*. A* $(k + k(k-1)/2) \times (k + k(k-1)/2)$ *matrix* $E[vv']$ *has full rank.*

Then for any $j = 1, \ldots, k$ *and* $t = 1, 2$, $E[\beta_{j,t}^2]$ *is identified.*

Full rank conditions (i) and (ii) do not require x_t to be continuously distributed. Therefore, one can test homogeneity of certain slope coefficients without the requirement that x_t is continuously distributed, as long as the support of (x_1, x_2) is rich enough.

Corollary 2.6. *Suppose that the assumptions of Theorem 2.5 hold. Then the covariates with homogeneous responses are identified.*

Proof. The response to a set of covariates $x_t^{(2)}$ is homogeneous if and only if $\text{Var}((x_{it}^{(2)})'\delta|x) = 0$. But

$$\text{Var}((x_{it}^{(2)})'\delta|x) = E[((x_{it}^{(2)})'\delta)^2|x] - \Big(E[(x_{it}^{(2)})'\delta|x]\Big)^2$$

Note that $E[(x_{it}^{(2)})'\delta|x]$ is identified (from a corresponding linear panel data model with fixed effects). Theorem 2.5 implies that $E[((x_{it}^{(2)})'\delta)^2|x]$ is identified. □

Remark 2.1. The result in Theorem 2.5 can be generalized to moments of order higher than 2, as long as covariates have rich enough support. Moreover, if the distribution of β_t is uniquely determined by its first M moments (where M is fixed and known), then with rich support (but not necessarily continuous covariates) we can identify the whole distribution of β_t. For the sketch of the proof of this result, please refer to the Appendix.

Identification of the covariates with homogeneous effects can serve as a way to reduce dimensionality of the vector of random coefficients β_t. Also, the identification result in

Theorem 2.3 requires continuity of Δx only for covariates with heterogeneous responses, since homogeneous responses are identified from a fixed effects regression.

2.2 Quantile Regression

A panel data quantile regression model is a special case of the random coefficients model:

$$y_{it} = x_{it}'\theta(u_{it}) + \alpha_i \tag{4}$$

where $u_{it} \sim U[0,1]$ is a uniformly distributed scalar random variables independent of x (and α). We assume that for any $x \in \mathcal{X}$ and $t = 1, 2$, the mapping $\tau \in (0,1) \mapsto x_t'\theta(\tau)$ is strictly increasing in τ. The goal is to identify the conditional quantiles of $x_{it}'\theta(u_{it})$ which take the following form:

$$Q_\tau(x_{it}'\theta(u_{it})|x_i) = x_{it}'\theta(\tau)$$

In what follows the subscript i is omitted, and the identification result for the quantile regression model (4) is given by the following theorem:

Theorem 2.7. *Suppose that assumptions A1 - A4 (or assumptions A1′ - A4′) are satisfied for* $\beta_t = \theta(u_t)$ *and* $t = 1, 2$. *Also, suppose that matrix* $E[x_t x_t']$ *has full rank. Then* $Q_\tau(x_{it}'\theta(u_{it})|x_i) = x_{it}'\theta(\tau)$ *is identified for any* $\tau \in (0,1)$.

Note that this result in does not necessarily require that x_t is continuously distributed on a unit ball. The only continuity assumption we need is the continuity of Δx (assumption A3′), and only in the case when we want to allow dependence between u_1, u_2 and α. The full rank condition of Theorem 2.7 is a standard full rank condition for linear quantile regression models.

3 Estimation

Both Theorem 2.1 and Theorem 2.3 provide a way to estimate the characteristic function of $x_t'\beta_t$. For continuously distributed covariates one can use kernel estimator to estimate the corresponding characteristic functions.

In particular, if we assume that β_1, β_2 and α are mutually independent conditional on x, Theorem 2.1 gives the following expression for the characteristic function of $x_t'\beta_t$:

$$\varphi_{x_t'\beta_t|x}(v) = \exp\left(\int_0^v \frac{\mathrm{i}E[y_t e^{\mathrm{i}s(y_t - y_{-t})}]}{E[e^{\mathrm{i}s(y_t - y_{-t})}]} ds - \mathrm{i}vE[\alpha|x]\right)$$

Given assumption A6 (random sampling), one can consistently estimate $\varphi_{x_t'\beta_t|x}(v)$ with

$$\hat{\varphi}_{x_t'\beta_t|x}(v) = \exp\left(\int_0^v \frac{\mathrm{i}\sum_{i=1}^n [y_{it} e^{\mathrm{i}s(y_{it} - y_{i,-t})}] K_h(x_i - x)}{\sum_{i=1}^n [e^{\mathrm{i}s(y_{it} - y_{i,-t})}] K_h(x_i - x)} ds - \mathrm{i}v\hat{E}[\alpha|x]\right)$$

where $K_h(\cdot)$ is a kernel function with bandwidth h and

$$\hat{E}[\alpha|x] = \frac{1}{2}\sum_{t=1}^{2}\left\{\frac{\sum_{i=1}^n y_{it} K_h(x_i - x)}{\sum_{i=1}^n K_h(x_i - x)} - x_t'\hat{\beta}_\mu\right\}$$

Here $x = (x_1', x_2')'$ and $\hat{\beta}_\mu$ is a standard fixed effects estimator of β_μ.

If one is willing to assume that only β_1 and β_2 are independent conditional on x, then according to Theorem 2.3 we have the following expression for the characteristic function of $x_1'\beta_1$:

$$\varphi_{x_1'\beta_1|x}(v) = \exp\left((1+h)\int_0^v \frac{\partial\varphi_{\Delta y|x}(\tilde{s}/(1+h); h)/\partial h}{\tilde{s}\varphi_{\Delta y|x}(\tilde{s}/(1+h); h)} d\tilde{s}\right)$$

Here one can consistently estimate $\varphi_{\Delta y|x}(s; h)$ with

$$\hat{\varphi}_{\Delta y|x}(s; h) = \frac{\sum_{i=1}^n [e^{\mathrm{i}s(y_{i2} - y_{i,1})}] K_{h_1}(x_{i1} - x_1) K_{h_2}(x_{i2} - (1+h)x_1)}{\sum_{i=1}^n K_{h_1}(x_{i1} - x_1) K_{h_2}(x_{i2} - (1+h)x_1)}$$

and under assumption A7$'$, the derivative w.r.t. h of $\hat{\varphi}_{\Delta y|x}(s; h)$ is a consistent estimator of the derivative of $\varphi_{\Delta y|x}(s; h)$, so that one can consistently estimate $\varphi_{x_1'\beta_1|x}(v)$ with

$$\hat{\varphi}_{x_1'\beta_1|x}(v) = \exp\left((1+h)\int_0^v \frac{\partial\hat{\varphi}_{\Delta y|x}(\tilde{s}/(1+h); h)/\partial h}{\tilde{s}\hat{\varphi}_{\Delta y|x}(\tilde{s}/(1+h); h)} d\tilde{s}\right)$$

Finally, Theorems 2.5 and 2.1 provide way to estimate (first), second, and higher non-central moments of the conditional distribution of $x_t'\beta_t$ under stationarity assumption A2$'$. For example, a second moment $m_2(x,t) = E[(x_t'\beta_t)^2|x]$ can be estimated in a linear regression of $(\Delta y)2 + 2(x_1'\hat{\beta}_\mu)(x_2'\hat{\beta}_\mu)$ on vector w defined as follows: let $w_{jm} = x_{j,1}x_{m,1} + x_{j,2}x_{m,2}$ for $j, m = 1, \ldots, k$, $m < j$, so that w is a $1 \times k + k(k-1)/2$ vector composed of w_{jm}.

3.1 Estimation of Conditional Quantiles

Previous section provides a way to construct a consistent estimator of a characteristic function (and hence of the cumulative distribution function) of $z_t'\beta_t$ under various sets of assumptions. Suppose now that we picked some consistent estimator $\hat{F}_{z_t|x}(\cdot)$ of the conditional distribution of $x_t'\beta_t$ conditional on x. With $\beta_t = \theta(u_t)$ and under monotonicity assumptions on function $x_t'\theta(\cdot)$, we have that

$$Q_\tau(x_t'\theta(u_t)|x) = x_t'\theta(\tau)$$

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 $z_t = x_t'\beta_t$ for each x_i in the sample, and then applying standard quantile regression method to the new "data" $\{(z_{it}^*, x_{it}) : i = 1, \ldots, R; t = 1, 2\}$. This procedure can be summarized by these two steps:

1. Let $\{(z_{it}^*, x_{it}) : i = 1, \ldots, R$ 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

$$\hat{\theta}(\tau) = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{R} \sum_{t=1}^{2} \rho_\tau \left(z_{it}^* - x_{it}'\theta\right) \tag{5}$$

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

As long as $R \to \infty$ as $n \to \infty$ and $\hat{F}_{z_t|x}(\cdot)$ is a consistent estimator of $F_{z_t|x}(\cdot)$ for any $x \in \mathcal{X}$, $\hat{\theta}(\tau)$ 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 $\hat{F}_{z_t|x}(\cdot)$.

4 Conclusion

This paper provides identification conditions for a class of random coefficients panel data models with fixed effects. Depending on the assumptions one is willing to make about the relationship between random slope coefficients and random intercepts (fixed effects), one can identify the distribution of certain linear combinations of random coefficients either with or without requiring that the difference in covariates between two time periods is continuously distributed on a unit ball with center proportionate to the first-period value of the covariates. Moreover, if all covariates are continuously distributed on a unit ball, the joint distribution of random slope coefficients is identified. The identification results are constructive and provide a direct way to estimate the conditional distribution of certain linear combinations of random slope coefficients.

The panel data quantile regression model can be treated as a special case of random coefficients model. Therefore, this paper offers a novel approach to the identification and estimation of the linear quantile regression panel data models with fixed effects that requires only two time periods.

Finally, a panel data model that is linear in random coefficients allows to identify any finite number of moments of marginal distributions of random slope coefficients when the covariates have rich enough support (which does not necessarily require that covariates are continuously distributed). It is shown that each non-centered moment of order $p = 1, 2 \dots, M$ can be estimated at a parametric rate with relatively mild distributional assumptions. This result can be used to test various distributional assumptions about random slope coefficients. In particular, one needs only first and second moments to test for homogeneity of the responses for a specified group of covariates.

References

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.

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.

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

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.

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

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

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.

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

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

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.

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

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

Graham, B. S., and J. L. Powell (2012). "Identification and Estimation of Average Partial Effects in 'Irregular' Correlated Random Coefficient Models," *Econometrica*, 80 (5), 2105-2152

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

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.

Hoderlein, S., J. Klemelä, and E. Mammen (2007). "Reconsidering the Random Coecient Model", working paper, Brown University.

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

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

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

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.

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

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

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

A Appendix

A.1 Proof of Theorem 2.1

First note that Assumption (stationarity and mean independence) implies that $E(x_t'\beta_t|x) = x_t'E(\beta_t) = x_t'\beta_\mu$, and vector β_μ is identified when matrix $E(\Delta x\Delta x')$ has full rank (this is a standard identification in linear panel data model with fixed effects). Therefore, $E(\alpha|x) = E(y_t|x) - x_t'\beta_\mu$ is identified.

Next, let $z_t = x_t'\beta_t$. Then following Evdokimov (2010), note that

$$\varphi_{y_1,y_2|x}(s_1,s_2) = \varphi_{z_1|x}(s_1)\varphi_{z_2|x}(s_2)\varphi_{\alpha|x}(s_1+s_2)$$

Therefore,

$$\begin{aligned}\frac{\partial\varphi_{y_1,y_2|x}(s_1,s_2)}{\partial s_1} =& \varphi'_{z_1|x}(s_1)\varphi_{z_2|x}(s_2)\varphi_{\alpha|x}(s_1+s_2)\\ &+ \varphi_{z_1|x}(s_1)\varphi_{z_2|x}(s_2)\varphi'_{\alpha|x}(s_1+s_2)\end{aligned}$$

The assumption (non-vanishing characteristic functions) implies that

$$\frac{\partial\varphi_{y_1,y_2|x}(s,-s)/\partial s_1}{\varphi_{y_1,y_2|x}(s,-s)} = \frac{\varphi'_{z_1|x}(s)}{\varphi_{z_1|x}(s)} + \frac{\varphi'_{\alpha|x}(0)}{\varphi_{\alpha|x}(0)}$$

and similarly,

$$\frac{\partial\varphi_{y_1,y_2|x}(-s,s)/\partial s_2}{\varphi_{y_1,y_2|x}(-s,s)} = \frac{\varphi'_{z_2|x}(s)}{\varphi_{z_2|x}(s)} + \frac{\varphi'_{\alpha|x}(0)}{\varphi_{\alpha|x}(0)}$$

Here $\varphi_{\alpha|x}(0) = 1$ and $\varphi'_{\alpha|x}(0) = iE[\alpha|x] = iE[y_t|x] - x_t'\beta_\mu$. So that solving for e.g. $\varphi_{z_1|x}(s)$ we have:

$$\varphi_{z_1|x}(v) = \exp\left(\int_0^v \frac{\partial\varphi_{y_1,y_2|x}(s,-s)/\partial s_1}{\varphi_{y_1,y_2|x}(s,-s)}ds - ivE[\alpha|x]\right)$$

so that $\varphi_{z_1|x}(\cdot)$ (and hence $f_{z_1|x}(\cdot)$ and $F_{z_1|x}(\cdot)$) are identified. For example,

$$F_{z_1|x}(u) = -\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-ivu}}{iv}\varphi_{z_1|x}(v)dv$$

$$= -\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\exp\left(-ivu + \int_0^v \frac{\partial\varphi_{y_1,y_2|x}(s,-s)/\partial s_1}{\varphi_{y_1,y_2|x}(s,-s)}ds - ivE[\alpha|x]\right)}{iv}dv$$

Finally,

$$\varphi_{\alpha|x}(v) = \frac{\varphi_{y_1,y_2|x}(v,0)}{\varphi_{z_1|x}(v)}$$

so that $f_{\alpha|x}(\cdot)$ is also identified. □

A.2 Proof of Theorem 2.3

Without loss of generality, let's fix x_1 and denote $z_t = x_t'\beta_t$. Consider $x_2 = (1+h)x_1$, so that $x = (x_1', (1+h)x_1')'$. Then assumptions A1′, A2′ and A3′ imply that for $\Delta y = y_2 - y_1$

$$\varphi_{\Delta y|x}(s;h) = \varphi_{z_2|x}(s)\varphi_{z_1|x}(-s)\varphi_{z_1|x}(s(1+h))\varphi_{z_1|x_1}(-s)$$

Then

$$\frac{\partial\varphi_{\Delta y|x}(s;h)}{\partial h} = \varphi'_{z_1|x}(s(1+h))\varphi_{z_1|x}(-s)s$$

so that

$$\frac{\partial\varphi_{\Delta y|x}(s;h)/\partial h}{\varphi_{\Delta y|x}(s;h)} = s\frac{\varphi'_{z_1|x}(s(1+h))}{\varphi_{z_1|x}(s(1+h))}$$

Stationarity assumption A2′ implies that $\frac{\partial\varphi_{\Delta y|x}(s)}{\partial h} = O(s^2)$ as $s \to 0$. Let $\tilde{s} = s(1+h)$. Then for some fixed $h \neq 1$ we have:

$$\frac{\varphi'_{z_1|x}(\tilde{s})}{\varphi_{z_1|x}(\tilde{s})} = \frac{\partial\varphi_{\Delta y|x}(\tilde{s}/(1+h);h)/\partial h}{\varphi_{\Delta y|x}(\tilde{s}/(1+h);h)}\frac{(1+h)}{\tilde{s}}$$

so that

$$\varphi_{z_1|x}(v) = \exp\left((1+h)\int_0^v \frac{\partial\varphi_{\Delta y|x}(\tilde{s}/(1+h);h)/\partial h}{\tilde{s}\varphi_{\Delta y|x}(\tilde{s}/(1+h);h)}d\tilde{s}\right)$$

Therefore, $\varphi_{z_1|x}(\cdot)$ is identified for any x. Similarly, one can show that $\varphi_{z_2|x}(\cdot)$ is identified for any x in the support, so that $f_{z_t|x}(\cdot)$ for $t = 1, 2$ are identified. $\square$

A.3 Proof of Theorem 2.5

Without loss of generality, assume that $x_t = (x'_{1,t}, x'_{2,t})'$, $\beta_t = (\beta'_{1,t}, \beta'_{2,t})'$ and that $\beta_{2,t} \equiv \beta_{2,f}$, where $\beta_{2,f}$ is a constant (vector). Here $E[x'_{2,t}\beta_{2,t}|x]$ is identified since $\beta_\mu = E[\beta_t|x]$ is identified. Let $m_1(x,t) = E[x'_t\beta_t|x]$. Then $m_1(x,t)$ is identified for any x and $t = 1, 2$.

Similarly, let $m_2(x,t) = E[(x'_t\beta_t)^2|x]$. Note that $m_2(x,t)$ is a homogeneous polynomial of degree 2 in $x_{1,t}, \ldots, x_{k,t}$. That is,

$$m_2(x,t) = \sum_{j=1}^{k} \sum_{m \le j}^{k} \gamma_{j,m} x_{j,t} x_{m,t}$$

where $\gamma_{j,m}$ are the (unknown) coefficients. Our goal is to show that coefficients $\{\gamma_{j,m}\}_{j,m=1}^{k}$ are identified.

Note that assumptions A1$'$ and A2$'$ imply that

$$\begin{aligned} E[(y_2 - y_1)^2|x] &= m_2(x,1) + m_2(x,2) - 2m_1(x,1)m_1(x,2) \\ &= \sum_{j=1}^{k} \sum_{m=1}^{k} \gamma_{j,m}(x_{j,1}x_{m,1} + x_{j,2}x_{m,2}) - 2m_1(x,1)m_1(x,2) \\ &= w'\gamma - 2m_1(x,1)m_1(x,2) \end{aligned}$$

where γ is a vector composed of γ_{jm}. Here $m_1(x,t)$ are identified for $t = 1, 2$. Therefore, since matrix $E[ww']$ has full rank, γ is identified, which implies that $m_2(x,t)$ is identified for $t = 1, 2$.

Finally, let $\sigma_j^2 = E[\beta_{j,t}^2]$ and $\sigma_{jm} = E[\beta_{j,t}\beta_{m,t}]$. Then

$$m_2(x,t) = \sum_{j=1}^{k} \sigma_j^2 x_{j,t}^2 + 2\sum_{j=1}^{k} \sum_{m<j} \sigma_{im} x_{j,t} x_{m,t} \tag{6}$$

Since matrix $E[vv']$ has full rank, σ_j^2 and σ_{jm} are identified for all $j = 1, \ldots, k$ and $m < j$. $\square$

A.4 Proof of Remark 2.1

For any p, let $m_p(x,t) = E[(x_t'\beta_t)^p|x]$. Then $m_p(x,t)$ is a homogeneous polynomial of degree p in $x_{1,t}, \ldots, x_{k,t}$. That is, it can be represented as

$$m_p(x,t) = \sum_{l_1+\ldots+l_k=p} \gamma_{l_1,\ldots,l_k}(k) x_{1,t}^{l_1} \ldots x_{k,t}^{l_k}. \tag{7}$$

That is, it is sufficient to show that under the coefficients $\{\gamma_{l_1,\ldots,l_d}(k), l_1+\ldots+l_k = p\}$ are identified for any $p \leq M$.

Since matrix $E[\Delta x \Delta x']$ has full rank, we can identify $m_1(x,t)$ (see e.g. the proof of Theorem 2.1).

Now suppose that for any $1 < q \leq p-1$ and any $x = \in \mathcal{X}$, $m_q(x,1)$ and $m_q(x,2)$ are identified. Since β_1 and β_2 are independent conditional on x, then for any p

$$E[(y_2 - y_1)^p|x] = \sum_{q=0}^{p} \binom{p}{q} (-1)^q m_{p-q}(x,1) m_q(x,2).$$

Therefore, we have:

$$m_p(x,1) + (-1)^p m_p(x,2) = E[(y_2-y_1)^p|x] - \sum_{q=1}^{p-1} \binom{p}{q} (-1)^q m_{p-q}(x,1) m_p(x,2) \tag{8}$$

The right-hand side of equation (8) is identified since $m_{p-q}(x_1)$ and $m_q(x_2)$ are assumed to be identified for any $x \in \mathcal{X}$ and any $1 \leq q \leq p-1$. The right-hand side of equation (8) is linear in vector $\Gamma(p)$ whose typical element is $\gamma_{l_1,\ldots,l_k}(p)$. That is,

$$m_p(x,1) + (-1)^k m_p(x,2) = w(p)'\Gamma(p),$$

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

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

With rich enough support, the matrix $E[w(p)w(p)']$ has full rank for $p \leq M$, and therefore vector $\Gamma(p)$ is identified. In other words, $\{\gamma_{l_1,\ldots,l_k}(k), l_1+\ldots+l_k = p\}$ are all identified, which immediately implies that $m_p(x,1)$ and $m_p(x,2)$ are also identified for any $x \in \mathcal{X}$. □

A.5 Proof of Theorem 2.7

Theorem 2.1 or Theorem 2.3 imply that for any x_t, the distribution of $x_t'\beta_t = x_t'\theta(u_t)$ is identified. The additional full rank condition of Theorem 2.7 ensures that the quantile slope coefficients for the conditional quantiles of this distribution are identified. $\square$