Quantile and Average Effects in Nonseparable Panel Models [1]

V. Chernozhukov, I. Fernandez-Val, W. Newey
MIT, Boston University, MIT

First Version, July 2009
Revised, September 2009

[1] We thank seminar participants of the SITE 2009 Workshop at Stanford and Bryan Graham for comments. We gratefully acknowledge research support from the NSF.

Abstract

This paper gives identification and estimation results for quantile and average effects in non-separable panel models, when the structural function and the distribution of period specific disturbances do not vary over time and regressors are discrete. We obtain bounds in static models (possibly with additive time effects) and in dynamic models. We derive rates at which the bounds tighten as the number T of time series observations grows.

1 Introduction

This paper gives identification and estimation results for quantile and average effects in nonseparable panel models, when the structural function and the distribution of period specific disturbances do not vary over time and regressors are discrete. We obtain bounds in static models (possibly with additive time effects) and in dynamic models. We derive rates at which the bounds tighten as the number T of time series observations grows. We also show in numerical calculations that the bounds may be very tight for small T, suggesting their usefulness in practice. We give an empirical illustration.

Nonseparable models are often needed to model important features of economic problems as discussed by Altonji and Matzkin (2005), Imbens and Newey (2009), and others. Also, Browning and Carro (2007) showed that economics motivates multiple sources of heterogeneity (not just an additive effect), and showed their importance in an application. Recently Hoderlein and White (2009) have considered a nonseparable panel data model that is close to the one we study.

Much of the work on nonseparable models in panel data (and other settings) has relied on control variables that arise from restricting the correlation between regressors and individual effects. Control variables are functions of observables such that the regressors and individual effects are independent conditional on those variables. Results on control variables for panel data are given by Chamberlain (1984), Altonji and Matzkin (2005), and Bester and Hansen (2009). We consider a different source of potential identification, time homogeneity of the structural function and of the distribution of idiosyncratic disturbances. Similar conditions have been used for identification by Chamberlain (1982), Manski (1987), Honore (1992), Hahn (2001), Wooldridge (2005), Chernozhukov, Fernandez-Val, Hahn, and Newey (2007), Graham and Powell (2008), and Hoderlein and White (2009), among others.

This paper is the first to consider identification of the quantile structural function (QSF) of Imbens and Newey (2009) and the average structural function (ASF) of Blundell and Powell (2003) under time homogeneity. We find that it is not possible to identify the QSF and ASF in panel data with discrete regressors though certain conditional effects may be identified. We give easily computed bounds for the QSF and ASF. We show that these bounds can be quite tight and can shrink exponentially fast as $T \longrightarrow \infty$, making the bounds potentially important in practice. We also allow for additive time specific effects or dynamics.

This paper is different than Honoré and Tamer (2006) and Chernozhukov, Hahn, and Newey (2004). Those papers derived bounds in semiparametric panel models where only individual location effects are present. This paper allows for slope effects also and considers nonparametric models.

In Section 2 we give the nonseparable models we consider and describe the QSF and ASF.

Section 3 derives bounds for the static case, conditional on an individual effect. Section 4 shows how additive time effects may be included. Section 5 gives bounds for the dynamic case. Section 6 considers consistency and rates as T grows.

2 The Model and Effects

The data consist of n observations $Y_i = (Y_{i1}, ..., Y_{iT})'$ and $X_i = [X_{i1}, ..., X_{iT}]'$, for a dependent variable Y_{it} and a vector of regressors X_{it}. We will assume throughout that (Y_i, X_i), $(i = 1, ..., n)$, are independent and identically distributed observations and that the support of X_i is finite (so X_{it} is discrete). A useful example is binary X_{it}, where $X_{it} \in \{0, 1\}$.

We consider a nonseparable model of the form

$$Y_{it} = g_0(X_{it}, \alpha_i, \varepsilon_{it}), (i = 1, ..., n; t = 1, ..., T), \tag{1}$$

where α_i and ε_{it} are unobserved disturbances that can have any dimension. The α_i is a vector of time invariant individual effects that often represents individual heterogeneity. The ε_{it} is a vector of period specific disturbances. Altonji and Matzkin (2005) considered this model.

By discreteness of X_{it} this model can also be written as a linear model with random coefficients. Suppose that X_{it} takes on the same J values $\{x_1, ..., x_J\}$ for each t and let D_{it} be a vector of dummy variables, $D_{itj} = 1(X_{it} = x_j)$. Let $\beta_j(\alpha_i, \varepsilon_{it}) = g(x_j, \alpha_i, \varepsilon_{it})$ and $\beta(\alpha_i, \varepsilon_{it}) = (\beta_1(\alpha_i, \varepsilon_{it}), ..., \beta_J(\alpha_i, \varepsilon_{it}))'$. Then equation (1) can also be written as

$$Y_{it} = D_{it}'\beta(\alpha_i, \varepsilon_{it}).$$

We consider identification in static and dynamic models under time homogeneity of the conditional distribution of ε_{it}. Time homogeneity in the static model is

$$\varepsilon_{it}|X_i, \alpha_i \overset{d}{=} \varepsilon_{is}|X_i, \alpha_i, \text{ for all } s, t. \tag{2}$$

This condition states that the conditional distribution of ε_{it} given X_i and α_i does not depend on t. This condition imposes conditional stationarity of the distribution of ε_{it} but allows for dependence of ε_{it} over time. An equivalent condition is $(\varepsilon_{it}, \alpha_i)|X_i \overset{d}{=} (\varepsilon_{is}, \alpha_i)|X_i$. The time invariant α_i has no distinct role in this model

Time homogeneity in the dynamic model is

$$\varepsilon_{it}|X_{i1}, ..., X_{it}, \alpha_i \overset{d}{=} \varepsilon_{i1}|X_{i1}, \alpha_i, \text{ for all } t. \tag{3}$$

Here we restrict the distribution of ε_{it} conditional on current and past X_{it} and α_i, requiring that it only depend on the first X_{i1} and α_i. In this model conditioning on α_i does play an important role, making ε_{it} independent of the regressor observations except for the first time period. This condition allows for dynamic feedback between ε_{it} and future X_{is} (i.e. with $s > t$). An important example is a dynamic binary choice model where Y_{it} is binary and $X_{it} = Y_{i,t-1}$.

We are here interested in two effects (functions) of X_{it} on the outcome, the average structural function (ASF) of Blundell and Powell (2003) and the quantile structural function (QSF) of Imbens and Newey (2009). The ASF is

$$\mu(x) = E[g_0(x, \alpha_i, \varepsilon_{it})] = \int g_0(x, \alpha, \varepsilon) F(d\alpha, d\varepsilon).$$

This object is useful for quantifying the effect of x on the mean of the outcome Y_{it}. In the treatment effects literature the average treatment effect of changing x from $\bar{x}$ to $\tilde{x}$ is

$$\mu(\tilde{x}) - \mu(\bar{x}).$$

The QSF is the λ^{th} quantile of $g(x, \alpha_i, \varepsilon_{it})$ as a function of x (and λ). To describe it, define the CDF of $g_0(x, \alpha_i, \varepsilon_{it})$ to be

$$G(y, x) = E[1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)].$$

Note that the time homogeneity assumptions imply that this function does not depend on t. The QSF is the inverse of this function

$$q(\lambda, x) = G^{-1}(\lambda, x).$$

In the treatment effects literature the λ^{th} quantile treatment effect of changing x from $\bar{x}$ to $\tilde{x}$ is

$$q(\lambda, \tilde{x}) - q(\lambda, \bar{x}),$$

as in Lehmann (1974).

A condition that is implicit in these objects is that the distribution of $(\varepsilon_{it}, \alpha_i)$ does not vary over time. This condition clearly holds in the static model and is implied by the dynamic one. Note that the distribution of ε_{it} given X_{i1}, α_i does not vary with i, implying the marginal distribution also does not vary with t either.

It may also be of interest to consider the ASF and QSF conditional on certain sequences X_i. For example, one could consider the effect of unions on income for those who ever belonged to a union. This is straightforward to do in the static case and also in the dynamic case for certain X_i.

Chamberlain (1982), Hahn (2001), Wooldridge (2005), and Chernozhukov et. al (2007) have considered nonseparable conditional mean models. Those models are implied by the ones here under additional conditions on conditional independence conditions for X_i and ε_{it}.

Theorem 1: *Suppose that equation (1) is satisfied,* $E[|Y_{it}|] < \infty$*, and* $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ *for all* x. Let $m_0(x, \alpha) = \int g_0(x, \alpha, \varepsilon)F(d\varepsilon|\alpha)$*. If equation (2) is satisfied and the conditional independence condition* $\varepsilon_{it}|X_i, \alpha_i \overset{d}{=} \varepsilon_{it}|\alpha_i$ *holds, then*

$$E[Y_{it}|X_i, \alpha_i] = m_0(X_{it}, \alpha_i).$$

If equation (3) is satisfied and the conditional independence condition $\varepsilon_{it}|X_{i1}, ..., X_{it}, \alpha_i \overset{d}{=} \varepsilon_{it}|\alpha_i$ *holds, then*

$$E[Y_{it}|X_{i1}, ..., X_{it}, \alpha_i] = m_0(X_{it}, \alpha_i).$$

Also, $\int m_0(x, \alpha)F(d\alpha) = \mu(x)$

Proof: Under the first conditional independence condition and equation (2),

$$\begin{aligned} E[Y_{it}|X_i, \alpha_i] &= E[g_0(X_{it}, \alpha_i, \varepsilon_{it})|X_i, \alpha_i] = \int g_0(X_{it}, \alpha_i, \varepsilon)F_{\varepsilon|X,\alpha}(d\varepsilon|X_i, \alpha_i) \\ &= \int g_0(X_{it}, \alpha_i, \varepsilon)F_{\varepsilon|\alpha}(d\varepsilon|\alpha_i) = m_0(X_{it}, \alpha_i). \end{aligned}$$

Similarly, under the second conditional independence condition and equation (3),

$$\begin{aligned} E[Y_{it}|X_{i1}, ..., X_{it}, \alpha_i] &= \int g_0(X_{it}, \alpha_i, \varepsilon)F_{\varepsilon|X,\alpha}(d\varepsilon|X_{i1}, ..., X_{it}, \alpha_i) \\ &= \int g_0(X_{it}, \alpha_i, \varepsilon)F_{\varepsilon|\alpha}(d\varepsilon|\alpha_i) = m_0(X_{it}, \alpha_i). \end{aligned}$$

The second conclusion follows by iterated expectations.Q.E.D.

A consequence of this is that the marginal effect, or average partial effect in the conditional mean sense is the same as the average treatment effect, i.e.

$$\int [m(\tilde{x}, \alpha) - m(\bar{x}, \alpha)]F(d\alpha) = \mu(\tilde{x}) - \mu(\bar{x}).$$

Without conditional independence of X_i and ε_i, the conditional means can depend on regressors in time periods other than t.

3 Bounds in the Static Model

In the static model there is a simple, fundamental result that provides information about the ASF. Let the support of X_i be $\{X^1, ..., X^K\}$. For all X^k such that $X^k_{t_k} = x$ for some t_k, we have

$$E[Y_{i,t_k}|X_i = X^k] = E[E[g_0(X_{it_k}, \alpha_i, \varepsilon_{it_k})|X_i = X^k, \alpha_i]|X_i = X^k] = E[g_0(x, \alpha_i, \varepsilon_{it})|X_i = X^k],$$

where the last equality follows by the time homogeneity conditions. That is, the ASF conditional on $X_i = X^k$ is equal to the expectation of Y_{it} for any t with $X_{it} = x$. This result generally does not suffice to identify the ASF because not all support points X^k have a time period with the regressor equal to x. When $g(x, \alpha_i, \varepsilon_{it})$ is bounded this does lead to bounds that can be quite tight even for small T. Also, under quite general conditions the probability of x not being a component of X_i shrinks to zero, leading to identification as $T \longrightarrow \infty$.

To describe the bounds, let $\mathcal{K}(x) = \{k : X^k_{t_k} = x \text{ for some } t_k\}$, $\bar{\mathcal{K}}(x)$ be the complement in $\{1, ..., K\}$, and $\mathcal{P}^k = \Pr(X_i = X^k)$. Define $\bar{\mathcal{P}}(x) = \sum_{k \in \bar{\mathcal{K}}(x)} \mathcal{P}^k$ to be the probability that x does not appear in any time period for X_i.

THEOREM 2: *If equations (1) and (2) are satisfied and $B_\ell \leq g(x, \alpha_i, \varepsilon_{it}) \leq B_u$ for constants B_ℓ and B_u and all x, then*

$$\mu_\ell(x) \leq \mu(x) \leq \mu_u(x),$$

where

$$\mu_\ell(x) = \sum_{k \in \mathcal{K}(x)} \mathcal{P}^k E[Y_{i,t_k} | X_i = X^k] + B_\ell \bar{\mathcal{P}}(x), \mu_u(x) = \mu_\ell(x) + \bar{\mathcal{P}}(x)(B_u - B_\ell).$$

Proof of Theorem 2: For $k \in \mathcal{K}(x)$ we have $X^k_{t_k} = x$, so that

$$E[Y_{i,t_k} | X_i = X^k] = E[g_0(X^k_{t_k}, \alpha_i, \varepsilon_{it_k}) | X_i = X^k] = E[g_0(x, \alpha_i, \varepsilon_{it}) | X_i = X^k].$$

For $k \in \bar{\mathcal{K}}(x)$ we have

$$B_\ell \leq E[g_0(x, \alpha_i, \varepsilon_{it}) | X_i = X^k] \leq B_u.$$

Multiplying by $\mathcal{P}^k$ and then adding over k gives the result. Q.E.D.

Corresponding bounds on treatment effect are then given by

$$\mu_\ell(\tilde{x}) - \mu_u(\bar{x}) \leq \mu(\tilde{x}) - \mu(\bar{x}) \leq \mu_u(\tilde{x}) - \mu_\ell(\bar{x}).$$

These bounds may be sharpened by imposing restrictions, such as monotonicity of treatment effects.

These bounds are the same as those derived for the marginal effect in a conditional mean model by Chernozhukov, Fernandez-Val, Hahn, and Newey (2007). Here we have shown that these bounds have a different interpretation as bounds on the ASF in the nonseparable model.

The bounds depend on the probability that none of the components of X_i is equal to x. For example, consider $X_{it} \in \{0, 1\}$. Suppose that $T = 2$. The support of X_i is $\{X^1, ..., X^4\}$, $X^1 = (0, 0)'$, $X^2 = (0, 1)'$, $X^3 = (1, 0)'$, $X^4 = (1, 1)'$. Let $x = 1$, so that $\mathcal{K}(1) = \{2, 3, 4\}$, $t_2 = 2$, $t_3 = 1$, and $t_4 = 1$ (or $t_4 = 2$). Also, $\bar{\mathcal{K}}(1) = \{1\}$, $\bar{\mathcal{P}}(1) = \Pr(X_i = (0, 0)')$, $\mu_\ell(1) = \sum_{k=2}^4 \mathcal{P}^k E[Y_{i,t_k} | X_i = X^k] + \mathcal{P}^1 B_\ell$ and $\mu_u(1) = \mu_\ell(1) + \mathcal{P}^1(B_u - B_\ell)$. Then the width of the bounds is $\mathcal{P}^1(B_u - B_\ell)$. For general T, the width is $\Pr(X_i = (0, ..., 0)')(B_u - B_\ell)$ that may decrease quickly as T grows.

Similarly to the treatment effects literature, we may be interested in the average structural function, or the average treatment effect, conditional on certain X_i values. For example, if $X_{it} \in \{0, 1\}$ represents treatment then we might be interested on the effect of treatment conditional on ever treated, i.e. conditional on $X_i \neq (0, ..., 0)'$. Tighter bounds for such effect can be formed and in some cases the effects may be identified.

The QSF bounds are obtained by replacing Y_{it} by $1(Y_{it} \leq y)$ in the ASF bounds, that is bounded below by 0 and above by 1, and inverting as a function of y. The bounds are based on the fundamental identification result that for any $k \in \mathcal{K}(x)$,

$$E[1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)|X_i = X^k] = E[1(Y_{i,t_k} \leq y)|X_i = X^k]$$

Bounds on the CDF $G(y, x)$ that are similar to Theorem 2 are

$$G_\ell(y, x) = \sum_{k \in \mathcal{K}(x)} \mathcal{P}^k E[1(Y_{i,t_k} \leq y)|X_i = X^k], \; G_u(y, x) = G_\ell(y, x) + \bar{\mathcal{P}}(x).$$

These can be inverted to give bounds on the QSF.

Theorem 3: *If equations (1) and (2) are satisfied then*

$$q_\ell(\lambda, x) \leq q(\lambda, x) \leq q_u(\lambda, x)$$

where

$$q_\ell(\lambda, x) = \begin{cases} -\infty, \lambda \leq \bar{\mathcal{P}}(x), \\ G_u^{-1}(\lambda, x), \lambda > \bar{\mathcal{P}}(x). \end{cases}, q_u(\lambda, x) = \begin{cases} G_\ell^{-1}(\lambda, x), \lambda < 1 - \bar{\mathcal{P}}(x), \\ +\infty, \lambda \geq 1 - \bar{\mathcal{P}}(x). \end{cases}.$$

Proof of Theorem 3: For $k \in \mathcal{K}(x)$ we have $X_{t_k}^k = x$, so that

$$\begin{aligned} E[1(Y_{i,t_k} &\leq y)|X_i = X^k] = E[1(g_0(X_t^k, \alpha_i, \varepsilon_{it_k}) \leq y)|X_i = X^k] \\ &= E[E[1(g_0(x, \alpha_i, \varepsilon_{it_k}) \leq y)|X_i, \alpha_i]|X_i = X^k] \\ &= E[E[1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)|X_i, \alpha_i]|X_i = X^k] \\ &= E[1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)|X_i = X^k]. \end{aligned}$$

For $k \in \bar{\mathcal{K}}(x)$ we have

$$0 \leq E[1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)|X_i = X^k] \leq 1.$$

Multiplying by $\mathcal{P}^k$ and adding up then gives

$$\begin{aligned} G_\ell(y, x) &= \sum_{k \in \mathcal{K}(x)} \mathcal{P}^k E[1(Y_{i,t_k} \leq y)|X_i = X^k] = \sum_{k \in \mathcal{K}(x)} \mathcal{P}^k E[1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)|X_i = X^k] \\ &\leq \sum_k \mathcal{P}^k E[1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)|X_i = X^k] = G(y, x) \leq G_\ell(y, x) + \bar{\mathcal{P}}(x) = G_u(y, x). \end{aligned}$$

The conclusion then follows by inverting. Q.E.D.

Bounds for quantile treatment effects can then be formed in the usual way as

$$q_\ell(\lambda, \tilde{x}) - q_u(\lambda, \bar{x}) \le q(\lambda, \tilde{x}) - q(\lambda, \bar{x}) \le q_u(\lambda, \tilde{x}) - q_\ell(\lambda, \bar{x}).$$

Estimation is straightforward. We can replace expectations by sample averages and the indicator function in the QSF bounds by a smoothed version, as in Yu and Jones (1998). When $X_{it} = x$ for multiple t we just average over the available time periods. This is not efficient but minimum distance would be difficult with small cells that will tend to happen when we are conditioning on every possible realization of $X_i = (X_{i1}, ..., X_{iT})'$. We can then use confidence intervals as in Chernozhukov, Hong, and Tamer (2007) or as in Beresteanu and Molinari (2008) based on joint asymptotic normality of the upper and lower bounds.

It is straightfoward to modify these bounds to apply to the ASF or QSF conditional on X_i being in $\mathcal{X} \subseteq \{X^1, ..., X^K\}$ of its support. This could be replacing the sums in the bounds with sums over support points in $\mathcal{X}$, divided by $\Pr(\mathcal{X})$. If x appears in some time period for every support point in $\mathcal{X}$ then the conditional ASF or QSF would be identified.

4 Additive Time Effects in Static Models

In static models it is possible to generalize the time homogeneity of $g_0(x, \alpha, \varepsilon)$ to allow for additive time effects. We consider a model where

$$Y_{it} = g_{t0}(X_{it}, \alpha_i, \varepsilon_{it}), \;\; g_{t0}(x, \alpha, \varepsilon) = \tau_t(x) + g_0(x, \alpha, \varepsilon), \;\; \tau_1(x) \equiv 0. \tag{4}$$

Graham and Powell (2008) consider a linear random coefficients model with additive time effects of this form. Here the ASF and QSF depend on t and are given by

$$\begin{aligned} \mu_t(x) &= \tau_t(x) + \int g_0(x, \alpha, \varepsilon) F(d\varepsilon, d\alpha), \\ q_t(\lambda, x) &= \lambda^{th} \text{ quantile of } \tau_t(x) + g_0(x, \alpha_i, \varepsilon_{it}) \\ &= \tau_t(x) + \lambda^{th} \text{ quantile of } g_0(x, \alpha_i, \varepsilon_{it}). \end{aligned}$$

Without dynamics one can easily identify such effects and allow for them in the estimation of the ASF and QSF.

To identify these effects let $\mathcal{X}_{tj}(x) = \{X : X_t = x, X_{t-j} = x\}$. Note that

$$\begin{aligned} E[Y_{it} - Y_{i,t-j} | X_i &\in \mathcal{X}_{tj}(x)] = \tau_t(x) - \tau_{t-j}(x) + E[g_0(x, \alpha_i, \varepsilon_{it}) - g_0(x, \alpha_i, \varepsilon_{i,t-j}) | X_i \in \mathcal{X}_{tj}(x)] \\ &= \tau_t(x) - \tau_{t-j}(x). \end{aligned}$$

Thus differences over time of the time effects are identified by corresponding differences of the Y_{it} conditional on x being constant over time. With the normalization $\tau_1(x) = 0$ it then follows that each $\tau_t(x)$ is identified as long as $\Pr(X_i \in \mathcal{X}_s(x)) > 0$ for each s where $\mathcal{X}_s(x) := \mathcal{X}_{s1}(x)$. Indeed, they appear to be overidentified for $T > 2$, since different conditioning sets can be used to recover that $\tau_t(x)$. Here we give an identification result that uses first differences.

Theorem 4: *If equations (2) and (4) are satisfied,* $E[|Y_{it}|] < \infty$ *for all* t*, and* $\Pr(X_i \in \mathcal{X}_s(x)) > 0$ *for each* $s = 2, ..., t$*, then*

$$\tau_t(x) = \sum_{s=2}^{t} E[Y_{is} - Y_{i,s-1} | X_i \in \mathcal{X}_s(x)].$$

Proof of Theorem 4: By $E[Y_{it} - Y_{i,t-1}|X_i \in \mathcal{X}_t(x)] = \tau_t(x) - \tau_{t-1}(x)$ and $\tau_1(x) = 0$ we have

$$\sum_{s=2}^{t} E[Y_{is} - Y_{i,s-1}|X_i \in \mathcal{X}_s(x)] = \sum_{s=2}^{t} [\tau_s(x) - \tau_{s-1}(x)] = \tau_t(x) - \tau_1(x) = \tau_t(x). Q.E.D.$$

To identify ASF and QSF proceed as above, first subtracting off $\tau_t(x)$ and then adding it back in.

Theorem 5: *If equations (2) and (4) are satisfied,* $\Pr(X_i \in \mathcal{X}_s(x)) > 0$ *for each* $s = 2, ..., t$*, and* $B_\ell \leq g_0(x, \alpha_i, \varepsilon_{it}) \leq B_u$ *for constants* B_ℓ *and* B_u *and all* x*, then*

$$\mu_{t\ell}(x) \leq \mu_t(x) \leq \mu_{tu}(x).$$

where

$$\begin{aligned} \mu_{t\ell}(x) &= \tau_t(x) + \sum_{k \in \mathcal{K}(x)} \mathcal{P}^k E[Y_{i,t_k} - \tau_{t_k}(x)|X_i = X^k] + B_\ell \bar{\mathcal{P}}(x), \\ \mu_{tu}(x) &= \mu_{t\ell}(x) + \bar{\mathcal{P}}(x)(B_u - B_\ell). \end{aligned}$$

Proof of Theorem 5: For $k \in \mathcal{K}(x)$ we have $X_{t_k}^k = x$, so that similarly to the proof of Theorem 3,

$$E[Y_{i,t_k} - \tau_{t_k}(x)|X_i = X^k] = E[g_0(X_{t_k}^k, \alpha_i, \varepsilon_{it_k})|X_i = X^k] = E[g_0(x, \alpha_i, \varepsilon_{it})|X_i = X^k].$$

For $k \in \bar{\mathcal{K}}(x)$ we have

$$B_\ell \leq E[g_0(x, \alpha_i, \varepsilon_{it})|X_i = X^k] \leq B_u.$$

The conclusion then follows by multiplying by $\mathcal{P}^k$, then adding over k, and adding back gives the result. Q.E.D.

To describe the quantile bounds redefine $G_\ell(y,x)$ as

$$G_\ell(y,x) = \sum_{k\in\mathcal{K}(x)} \mathcal{P}^k E[1(Y_{i,t_k} - \tau_{t_k}(x) \le y)|X_i = X^k], G_u(y,x) = G_\ell(y,x) + \bar{\mathcal{P}}(x).$$

THEOREM 6: *If equations (1) and (2) are satisfied then*

$$q_{t\ell}(\lambda,x) \le q_t(\lambda,x) \le q_{tu}(\lambda,x)$$

where

$$\begin{aligned} q_{t\ell}(\lambda,x) &= \begin{cases} -\infty, \lambda \le \bar{\mathcal{P}}(x), \\ \tau_t(x) + G_u^{-1}(\lambda,x), \lambda > \bar{\mathcal{P}}(x). \end{cases}, \\ q_{tu}(\lambda,x) &= \begin{cases} \tau_t(x) + G_\ell^{-1}(\lambda,x), \lambda < 1-\bar{\mathcal{P}}(x), \\ +\infty, \lambda \ge 1-\bar{\mathcal{P}}(x). \end{cases}. \end{aligned}$$

Proof of Theorem 6: For $k \in \mathcal{K}(x)$ we have $X_{t_k}^k = x$, so that

$$\begin{aligned} E[1(Y_{i,t_k} - \tau_{t_k}(x) &\le y)|X_i = X^k] = E[1(g_0(X_{t_k}^k, \alpha_i, \varepsilon_{it_k}) \le y)|X_i = X^k] \\ &= E[1(g_0(x,\alpha_i,\varepsilon_{it}) \le y)|X_i = X^k]. \end{aligned}$$

For $k \in \bar{\mathcal{K}}(x)$ we have

$$0 \le E[1(g_0(x,\alpha_i,\varepsilon_{it}) \le y)|X_i = X^k] \le 1.$$

Multiplying by $\mathcal{P}^k$ and adding up then gives

$$\begin{aligned} G_\ell(y,x) &= \sum_{k\in\mathcal{K}(x)} \mathcal{P}^k E[1(Y_{i,t_k} - \tau_{t_k}(x) \le y)|X_i = X^k] \\ &= \sum_{k\in\mathcal{K}(x)} \mathcal{P}^k E[1(g_0(x,\alpha_i,\varepsilon_{it}) \le y)|X_i = X^k] \\ &\le G(y,x) \le G_u(y,x) + \bar{\mathcal{P}}(x). \end{aligned}$$

The conclusion then follows by inverting and adding back $\tau_t(x)$. Q.E.D.

The QSF bounds are unusual in that the trend terms in quantiles are identified from expectations. This approach depends crucially on $\tau_t(x)$ being constant (i.e. nonrandom). The ASF bounds will also apply when $\tau_t(x)$ has some forms of randomness, but the QSF bounds will not.

5 Bounds in the Dynamic Model

In static models we developed bounds using the partition of the support of X_i into all its separate points. In the dynamic case such a partition no longer works, because the disturbance ε_{it} is not time homogeneous conditional on X_{is} for all time periods. However we can partition the support of X_i differently to still obtain bounds.

To describe the bounds, consider the partition

$$\begin{aligned}\mathcal{X}_t(x) &= \{X : X_t = x,\ X_s \neq x\ \forall s < t\},\\ \bar{\mathcal{X}}(x) &= \{X : X_t \neq x\ \forall t\}.\end{aligned}$$

THEOREM 7: *If equations (1) and (3) are satisfied and for all x, and $B_\ell \leq g(x, \alpha_i, \varepsilon_{it}) \leq B_u$ for constants B_ℓ and B_u and all x, then*

$$\mu_\ell(x) \leq \mu(x) \leq \mu_u(x).$$

where

$$\mu_\ell(x) = \sum_{t=1}^{T} E[1(X_i \in \mathcal{X}_t(x))Y_{it}] + B_\ell \bar{\mathcal{P}}(x), \mu_u(x) = \mu_\ell(x) + \bar{\mathcal{P}}(x)(B_u - B_\ell).$$

Proof of Theorem 7: Since $X_i \in \mathcal{X}_t(x)$ only restricts $X_{it}, ..., X_{i1}$ and $X_{it} = x$ for any $X_i \in \mathcal{X}_t(x)$ we have

$$\begin{aligned}E[Y_{it}|X_i &\in \mathcal{X}_t(x)] = E[E[g_0(X_{it}, \alpha_i, \varepsilon_{it})|X_{it}, ..., X_{i1}, \alpha_i]|X_i \in \mathcal{X}_t(x)]\\ &= E[E[g_0(x, \alpha_i, \varepsilon_{it})|X_{it}, ..., X_{i1}, \alpha_i]|X_i \in \mathcal{X}_t(x)]\\ &= E[E[g_0(x, \alpha_i, \varepsilon_{i1})|X_{i1}, \alpha_i]|X_i \in \mathcal{X}_t(x)]\\ &= E[g_0(x, \alpha_i, \varepsilon_{i1})|X_i \in \mathcal{X}_t(x)].\end{aligned}$$

We also have

$$B_\ell \leq E[g_0(x, \alpha_i, \varepsilon_{i1})|X_i \in \bar{\mathcal{X}}(x)] \leq B_u.$$

Note that $\Pr(X_i \in \bar{\mathcal{X}}(x)) = \bar{\mathcal{P}}(x)$. Then adding up gives

$$\mu_\ell(x) \leq E[g_0(x, \alpha_i, \varepsilon_{i1})] \leq \mu_u(x).$$

since $\mathcal{X}_1(x), ..., \mathcal{X}_T(x), \bar{\mathcal{X}}(x)$ is a partition of the support of X_i. Note also that eq. Q.E.D.

An important example is the binary $Y_{it} \in \{0, 1\}$ case where $X_{it} = Y_{i,t-1}$. In this case $B_\ell = 0$, $B_u = 1$. Here $\bar{\mathcal{P}}(0) = \Pr(X_i = (1, ..., 1)')$ and $\bar{\mathcal{P}}(1) = \Pr(X_i = (0, ..., 0)')$. The bounds for $\mu(0)$ and $\mu(1)$ will be

$$\begin{aligned}\sum_{t=1}^{T} E[1(X_i &\in \mathcal{X}_t(0))Y_{it}] = \mu_\ell(0) \leq \mu(0) \leq \mu_u(0) = \mu_\ell(0) + \bar{\mathcal{P}}(0),\\ \sum_{t=1}^{T} E[1(X_i &\in \mathcal{X}_t(1))Y_{it}] = \mu_\ell(1) \leq \mu(1) \leq \mu_u(1) = \mu_\ell(1) + \bar{\mathcal{P}}(1).\end{aligned}$$

Then for $\delta = \sum_{t=1}^{T} E[\{1(X_i \in \mathcal{X}_t(1)) - 1(X_i \in \mathcal{X}_t(0))\}Y_{it}]$ we have

$$\delta - \bar{\mathcal{P}}(1) \leq \mu(1) - \mu(0) \leq \delta + \bar{\mathcal{P}}(0).$$

Width of the bounds is $\Pr(X_i = (1, ..., 1)') + \Pr(X_i = (0, ..., 0)')$. Informative in long panels but not short ones. This is a bounds solution to the problem of identifying duration dependence in the presence of unobserved heterogeneity (Feller, 1943, and Heckman, 1981). Note that

$$\mu(1) - \mu(0) = \int [\Pr(Y_{it} = 1 | Y_{i,t-1} = 1, \alpha) - \Pr(Y_{it} = 1 | Y_{i,t-1} = 0, \alpha)] F(d\alpha),$$

the effect of lagged Y_{it}, holding α_i fixed, averaged over α_i.

For bounds for the QSF define

$$\tilde{G}_\ell(y, x) = \sum_{t=1}^{T} E[1(X_i \in \mathcal{X}_t(x)) 1(Y_{it} \leq y)], \tilde{G}_u(y, x) = \tilde{G}_\ell(y, x) + \bar{\mathcal{P}}(x).$$

THEOREM 8: *If equations (1) and (3) are satisfied then*

$$q_\ell(\lambda, x) \leq q(\lambda, x) \leq q_u(\lambda, x)$$

where

$$q_\ell(\lambda, x) = \left\{ \begin{array}{c} -\infty, \lambda \leq \bar{\mathcal{P}}(x), \\ \tilde{G}_u^{-1}(\lambda, x), \lambda > \bar{\mathcal{P}}(x). \end{array} \right. , q_u(\lambda, x) = \left\{ \begin{array}{c} \tilde{G}_\ell^{-1}(\lambda, x), \lambda < 1 - \bar{\mathcal{P}}(x), \\ +\infty, \lambda \geq 1 - \bar{\mathcal{P}}(x). \end{array} \right. .$$

Proof of Theorem 8: Since $X_i \in \mathcal{X}_t(x)$ only restricts $X_{it}, ..., X_{i1}$ we have

$$\begin{aligned} E[1(Y_{it} &\leq y) | X_i \in \mathcal{X}_t(x)] = E[E[1(Y_{it} \leq y) | X_{it}, ..., X_{i1}, \alpha_i] | X_i \in \mathcal{X}_t(x)] \\ &= E[1(g_0(x, \alpha_i, \varepsilon_{i1}) \leq y) | \alpha_i] | X_i \in \mathcal{X}_t(x)]. \end{aligned}$$

We also have

$$0 \leq E[E[1(g_0(x, \alpha_i, \varepsilon_{i1}) \leq y) | \alpha_i] | X_i \in \bar{\mathcal{X}}(x)] \leq 1.$$

Then adding up and inverting gives the result. *Q.E.D.*

The bounds for the dynamic model also apply to the static model but there are advantages to condition on the entire sequence of X_i when there are no dynamics. One advantage is that it is possible to account for time effects as described above and, as far as we know, it is not possible to do this in the dynamic case. Also, it is straightforward to impose monotonicity in $g_0(x, \alpha_i, \varepsilon_{it})$ in the static case, and it is not known how to do this in the dynamic model.

6 Identification and Rates as $T \longrightarrow \infty$

The size of the bounds all depend on $\bar{\mathcal{P}}(x)$, the probability that x does not appear for any time period. Identification will be attained as $T \longrightarrow \infty$ if $\bar{\mathcal{P}}(x) \longrightarrow 0$. This convergence will occur under fairly weak conditions.

Theorem 9: *Suppose that equations (1) and (3) are satisfied, $\overrightarrow{X}_i = (X_{i1}, X_{i2}, ...)$ is stationary and, conditional on α_i, the support of each X_{it} is the marginal support of X_{it} and $\overrightarrow{X}_i$ is ergodic. If $B_\ell \leq g(x, \alpha_i, \varepsilon_{it}) \leq B_u$ for constants B_ℓ and B_u and all x, then $\mu_\ell(x) \longrightarrow \mu(x)$ and $\mu_u(x) \longrightarrow \mu_0(x)$ as $T \longrightarrow \infty$. If $0 < \lambda < 1$ and $G(y, x)$ is continuous and strictly monotonic in y on $\{y : 0 < G(y, x) < 1\}$ then $q_\ell(\lambda, x) \longrightarrow q(\lambda, x)$ and $q_u(\lambda, x) \longrightarrow q(\lambda, x)$ as $T \longrightarrow \infty$.*

Proof of Theorem 9: Let $Z_{iT} = \sum_{t=1}^{T} 1(X_{it} = x)/T$ Note that if $Z_{iT} > 0$ then $1(A_{iT}) = 1$ for the event A_{iT} that there exists $\tilde{t}$ such that $X_{i\tilde{t}} = x$. By the ergodic theorem, conditional on α_i we have $Z_{iT} \xrightarrow{as} \Pr(X_{it} = x \mid \alpha_i) > 0$ by the conditional support being equal to the marginal support. Therefore $\Pr(A_{iT} \mid \alpha_i) \geq \Pr(Z_{iT} > 0 \mid \alpha_i) \longrightarrow 1$ for almost all α_i. It then follows by the dominated convergence theorem that

$$\Pr(A_{iT}) = E[\Pr(A_{iT} \mid \alpha_i)] \longrightarrow 1.$$

Also note that $\Pr(A_{iT}) = 1 - \bar{\mathcal{P}}(x)$, so that

$$\bar{\mathcal{P}}(x) \longrightarrow 0.$$

The first conclusion then follows by Theorem 7.

Next, for notational convenience, suppress the x argument. It follows as previously with $1(g(x, \alpha_i, \varepsilon_{it}) \leq y)$ replacing Y_{it} that for all y, as $T \longrightarrow \infty$

$$G_u(y) - G_\ell(y) \leq \bar{\mathcal{P}} \longrightarrow 0.$$

Consider any $0 < \lambda < 1$. Let T be large enough so that $\lambda < 1 - \bar{\mathcal{P}}$. Then $q_u(\lambda)$ is finite and $G_\ell(q_u(\lambda)) = \lambda = G(q(\lambda))$. It follows by $q_u(\lambda) \geq q(\lambda)$ that

$$0 \leq G(q_u(\lambda)) - G(q(\lambda)) = G(q_u(\lambda)) - G_\ell(q_u(\lambda)) \leq \bar{\mathcal{P}} \longrightarrow 0.$$

Since $G(y)$ is strictly monotonic in a neighborhood of $q(\lambda)$ and $q_u(\lambda) \geq q(\lambda)$, it follows that $q_u(\lambda) \longrightarrow q(\lambda)$. An analogous argument shows that $q_\ell(\lambda) \longrightarrow q(\lambda).Q.E.D.$

This result gives conditions for identification as T grows, generalizing a result of Chamberlain (1982) for binary X_{it}. In addition, it shows that the bounds derived above shrink to the average and quantile effects as T grows. To explain when this identification would not hold it is helpful to consider a simple example where X_i is i.i.d. conditional on α_i. In that case

$$\bar{\mathcal{P}}(x) = E[\Pr(X_{it} \neq x|\alpha_i)^T].$$

This will not go to zero if and only if $\Pr(\Pr(X_{it} \neq x|\alpha_i) = 1) > 0$, that is, if and only if for some α_i the distribution of X_{it} puts zero weight on x. The marginal support being equal to the conditional support is the hypothesis that rules this out.

The rate at which the bounds converge in the general model is a complicated question. We can give a simple result if the conditional probability for $X_{it} = x$ is bounded away from zero.

Theorem 10: *Suppose that equations (1) and (3) are satisfied, $\overrightarrow{X_i}$ is stationary and Markov of order J conditional on α_i, and for some $\varepsilon > 0$,*

$$\Pr(X_{it} = x|X_{i,t-1}, ..., X_{i,t-J}, \alpha_i) \geq \varepsilon.$$

Then if $B_\ell \leq g(x, \alpha_i, \varepsilon_{it}) \leq B_u$,

$$\mu_u(x) - \mu_\ell(x) \leq (B_u - B_\ell)(1-\varepsilon)^{T-J}.$$

Also, if $0 < \lambda < 1$ *and* $G(y, x)$ *is continuously differentiable on a neighborhood of* $y = q(\lambda, x)$ *with a derivative bounded below by* $D_x > 0$, *then for a large enough* T

$$q_u(\lambda, x) - q_\ell(\lambda, x) \leq 2D_x^{-1}(1-\varepsilon)^{T-J}.$$

Proof of Theorem 10: Let $\Pi_{t=1}^T 1(X_{it} \neq x)$ be the indicator function for the event that none of the elements of X_i is equal to x so that $\bar{\mathcal{P}}(x) = E[\Pi_{t=1}^T 1(X_{it} \neq x)]$. By iterated expectations, for $T > J$,

$$\begin{aligned}
\bar{\mathcal{P}}(x) &= E[E[\Pi_{t=1}^T 1(X_{it} \neq x)] = E[\Pi_{t=1}^{T-1} 1(X_{it} \neq x) E[1(X_{iT} \neq x|X_{i,T-1}, ..., X_{i1}, \alpha_i]] \\
&= E[\{\Pi_{t=1}^{T-1} 1(X_{it} \neq x)\} \Pr(X_{iT} \neq x|X_{i,T-1}, ..., X_{i,T-J}, \alpha_i)] \leq (1-\varepsilon) E[\Pi_{t=1}^{T-1} 1(X_{it} \neq x)].
\end{aligned}$$

Repeating the argument for $T-1, ..., J$ gives

$$\bar{\mathcal{P}}(x) \leq (1-\varepsilon)^{T-J} E[\Pi_{t=1}^{J-1} 1(X_{it} \neq x)] \leq (1-\varepsilon)^{T-J}.$$

The first conclusion then follows by Theorem 7.

Next suppress the x argument and proceed as in the proof of Theorem 9. Note that $G'(y) > D_x$ for all y in a neighborhood of $q(\lambda)$ and that $G(q_u(\lambda)) - G(q(\lambda)) \leq \bar{\mathcal{P}}$ for large enough T. Using these and previous bounds and a mean value expansion gives

$$(1-\varepsilon)^{T-J} \geq \bar{\mathcal{P}} \geq G(q_u(\lambda)) - G(q(\lambda)) = G'(\bar{q}(\lambda))[q_u(\lambda) - q(\lambda)] \geq D_x[q_u(\lambda) - q(\lambda)] \geq 0,$$

where $\bar{q}(\lambda)$ lies between $q_u(\lambda)$ and $q(\lambda)$. Dividing by D_x then gives

$$D_x^{-1}(1-\varepsilon)^{T-J} \geq q_u(\lambda) - q(\lambda) \geq 0.$$

An analogous argument gives $D_x^{-1}(1-\varepsilon)^{T-J} \geq q(\lambda) - q_\ell(\lambda)$, so adding these inequalities gives the second conclusion. *Q.E.D.*

This result shows that the rate of convergence of the bounds will be exponential when the conditional probability that $X_{it} = x$ is bounded away from zero. The i.i.d. example can be used to illustrate what other kinds of results might occur. As discussed above, $\bar{\mathcal{P}}(x) = E[\Pr(X_{it} \neq x|\alpha_i)^T]$, so the rate of shrinkage depends on the thickness of the tails of the distribution of $\Pr(X_{it} \neq x|\alpha_i)$. If too much weight is put on conditional probabilities near one then the convergence may be slow. For example, suppose $X_{it} = 1(\alpha_i - v_{it} > 0)$, $\alpha_i \sim N(0,1)$, $v_{it} \sim N(0,1)$. Then

$$\bar{\mathcal{P}}(0) = E[\Phi(\alpha_i)^T] = \int \Phi(\alpha)^T \phi(\alpha) d\alpha = \left. \frac{\Phi(\alpha)^{T+1}}{T+1} \right|_{-\infty}^{+\infty} = \frac{1}{T+1},$$

which shrinks slower than exponentially. On the other hand, if α_i has any distribution with a compact support, Theorem 10 implies that the bounds shrink exponentially fast in T.

References

[1] Altonji, J.G. and R.L. Matzkin (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica* 73, 1053-1102.

[2] Beresteanu, A., and Molinari, F. (2008), "Asymptotic properties for a class of partially identified models," *Econometrica* 76, 763–814.

[3] Bester, C.A., and C. Hansen (2009), "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Economic and Business Statistics,* forthcoming.

[4] Blundell, R. and J.L. Powell (2003), "Endogeneity in Nonparametric and Semiparametric Regression Models," in M. Dewatripont, L. P. Hansen and S. J. Turnsovsky (eds.) *Advances in Economics and Econometrics*, Cambridge: Cambridge University Press.

[5] Browning, M. and J. Carro (2007), "Heterogeneity and Microeconometrics Modeling," in R. Blundell, W. Newey, T. Persson (eds.), *Advances in Theory and Econometrics, Vol. 3*, Cambridge, England: Cambridge University Press.

[6] Carro, J. M. (2007), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," *Journal of Econometrics* 140, 503-528.

[7] Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics* 18, 5–46.

[8] Chamberlain, G. (1984), "Panel Data," in Z. Griliches and M. Intriligator (eds), *Handbook of Econometrics*, Vol. 2, Amsterdam: North-Holland.MIT.

[9] Chernozhukov, V., J.Hahn, and W.K.Newey (2004), "Bound Analysis in Panel Models with Correlated Random Effects," working paper, MIT.

[10] Chernozhukov, V., I. Fernandez-Val, J.Hahn, and W.K.Newey (2007), "Identification and Estimation of Marginal Effects in Nonlinear Panel Models," working paper, MIT.

[11] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models,"*Econometrica* 75, 1243–1284.

[12] Feller, W. (1943), "On a General Class of Contagious Distributions," *Annals of Statistics* 14, 389-400.

[13] Fernandez-Val, I. (2008), "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models," working paper, Boston University.

[14] Graham, B.W. and J.L. Powell (2008), "Semiparametric Identification and Estimation of Correlated Random Coefficient Models for Panel Data," working paper, Boston Univ..

[15] Hahn, J. (2001), "Comment: Binary Regressors in Nonlinear Panel-Data Models with Fixed Effects," *Journal of Business and Economic Statistics* 19, 16-17.

[16] Heckman,J.J. (1981), "Statistical Models for Discrete Panel Data," in Manski, C.F. and D. McFadden eds., *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA.

[17] Hoderlein, S. and H. White (2009), "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects", working paper, Brown University.

[18] Honore, B.E. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica* 60, 533-565.

[19] Honore, B.E., and E. Tamer (2006), "Bounds on Parameters in Dynamic Discrete Choice Models", *Econometrica* 74, 611-629.

[20] Imbens, G.W. and W.K. Newey (2009) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica,* forthcoming.

[21] Lehmann, E. (1974) *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

[22] Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models From Binary Response Data," *Econometrica* 55, 357-362.

[23] Wooldridge, J.M. (2005): "Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models," *Review of Economics and Statistics* 87, 385–390.

[24] Yu, K. and M.C. Jones (1998): "Local Linear Quantile Regression," *Journal of the American Statistical Association* 93, 228-237.