# 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, October 2009

[1] We thank seminar participants at Stanford and Harvard-MIT econometrics workshops, B. Graham, and J. Hausman for comments. We gratefully acknowledge research support from the NSF.

**Abstract**

This paper gives identification and estimation results for quantile and average effects in nonseparable panel models, when the distribution of period specific disturbances does not vary over time. Bounds are given for interesting effects with discrete regressors that are strictly exogenous or predetermined. We allow for location and scale time effects and show how monotonicity can be used to shrink the bounds. We derive rates at which the bounds tighten as the number $T$ of time series observations grows and give an empirical illustration.

# 1 Introduction

This paper gives identification and estimation results for quantile and average effects in nonseparable panel models, when the distribution of period specific disturbances does not vary over time. Bounds are given for interesting effects with discrete regressors that are strictly exogenous or predetermined. We allow for location and scale time effects and show how monotonicity can be used to shrink the bounds. We derive rates at which the bounds tighten as the number $T$ of time series observations grows and give an empirical illustration.

Nonseparable models are often needed to model important features of economic problems as discussed by Altonji and Matzkin (2005) 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. 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 location and scale time effects or dynamics, and show how monotonicity can be used to tighten the bounds. The empirical illustration is based on Chamberlain's (1982) union wage effects application.

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, with regressors that are strictly exogenous conditional on an individual effect. Section 4 shows how location and scale time effects may be included. Section 5 gives bounds for the dynamic case with predetermined regressors. Section 6 gives bounds under monotonicity. Section 7 considers consistency and rates as $T$ grows. Section 8 gives the empirical example.

## 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.

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

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

$$\varepsilon_{it}|X_i, \alpha_i \stackrel{d}{=} \varepsilon_{i1}|X_i, \alpha_i, \text{ for all } t. \tag{2}$$

This condition states that the conditional distribution of $\varepsilon_{it}$ given $X_i$ and $\alpha_i$ does not depend on $t$. This condition imposes conditional stationarity of the distribution of $\varepsilon_{it}$ but allows for dependence of $\varepsilon_{it}$ over time.

An equivalent condition is $\tilde{\varepsilon}_{it}|X_i \stackrel{d}{=} \tilde{\varepsilon}_{i1}|X_i$ for $\tilde{\varepsilon}_{it} = (\alpha_i, \varepsilon_{it})$. The time invariant $\alpha_i$ has no distinct role in this model. The condition is just that whatever the disturbances are, their conditional distribution given $X_i$ does not depend on $t$. This seems a basic "ceteris paribus" assumption for panel data that amounts to the time period being "randomly assigned." In a linear model this condition is observationally equivalent to a more standard one involving an individual effect. Suppose that

$$Y_{it} = X_{it}'\beta + \alpha_i + \varepsilon_{it} = X_{it}'\beta + \tilde{\varepsilon}_{it}.$$

A linear model version of the time homogeneity condition is $E^*(\tilde{\varepsilon}_{it}|X_i) = E^*(\tilde{\varepsilon}_{i1}|X_i)$ for all $t$, where $E^*(\cdot|X_i)$ denotes linear projection on $X_i$. Then

$$\begin{aligned} E^*(Y_{it}|X_i) &= X_{it}'\beta + E^*(\tilde{\varepsilon}_{it}|X_i) = X_{it}'\beta + E^*(\tilde{\varepsilon}_{i1}|X_i) \\ &= X_{it}'\beta + \tilde{\alpha}_i, \tilde{\alpha}_i = E^*(\tilde{\varepsilon}_{i1}|X_i). \end{aligned}$$

This is same multivariate regression (Chamberlain, 1982) implied by an additive individual effect. Thus, in the linear model the time homogeneity condition is observationally equivalent to an additive individual effect.

The dynamic time homogeneity condition we impose is

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

Here we restrict the distribution of $\varepsilon_{it}$ conditional on current and past $X_{it}$ and $\alpha_i$, requiring that it only depends on $X_{i1}$ and $\alpha_i$. In this model conditioning on $\alpha_i$ does play an important role, making $\varepsilon_{it}$ independent of the regressor observations except for the first time period. This condition allows for dynamic feedback between $\varepsilon_{it}$ 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 $\lambda^{th}$ quantile of $g(x, \alpha_i, \varepsilon_{it})$ as a function of $x$ (and $\lambda$). 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 $\lambda^{th}$ 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 conditional distribution of $\varepsilon_{it}$ given $X_{i1}, \alpha_i$ does not vary with $t$, implying the marginal distribution of $(\varepsilon_{it}, \alpha_i)$ also does not vary with $t$.

Chamberlain (1982), Hahn (2001), Wooldridge (2005), and Chernozhukov et. al (2007) have considered nonseparable conditional mean models where the object of interest is an average partial effect. The nonseparable models given here imply those models with average treatment effect equal to the average partial effect.

THEOREM 1: *Suppose that equation (1) is satisfied, $E[|Y_{it}|] < \infty$, and $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ for all $x$. If equation (2) is satisfied then for $\tilde{\alpha} = X$ and $m_0(x, \tilde{\alpha}) = \int g_0(x, \alpha, \varepsilon) F(d\alpha, d\varepsilon|\tilde{\alpha})$.*

$$E[Y_{it}|X_i, \tilde{\alpha}_i] = m_0(X_{it}, \tilde{\alpha}_i), \mu(x) = \int m_0(x, \tilde{\alpha}) F(d\tilde{\alpha}).$$

*If equation (3) is satisfied then for $\tilde{\alpha} = (\alpha, X_1)$ and $m_0(x, \tilde{\alpha}) = \int g_0(x, \alpha, \varepsilon) F(d\varepsilon|\tilde{\alpha})$,*

$$E[Y_{it}|X_{it}, ..., X_{i1}, \tilde{\alpha}_i] = m_0(X_{it}, \tilde{\alpha}_i), \mu(x) = \int m_0(x, \tilde{\alpha}) F(d\tilde{\alpha}).$$

Proof of Theorem 1: Under equation (2), for $\tilde{\alpha} = X$,

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

Similarly, under equation (3), for $\tilde{\alpha}_i = (\alpha_i, X_{1i})$,

$$\begin{aligned} E[Y_{it}|X_{it}, ..., X_{i1}, \tilde{\alpha}_i] &= \int g_0(X_{it}, \alpha_i, \varepsilon) F(d\varepsilon|X_{it}, ..., X_{i1}, \alpha_i) \\ &= \int g_0(X_{it}, \alpha_i, \varepsilon) F(d\varepsilon|\tilde{\alpha}_i) = m_0(X_{it}, \tilde{\alpha}_i), \\ \int m_0(x, \tilde{\alpha}) F(d\tilde{\alpha}) &= \int g_0(x, \alpha, \varepsilon) F(d\varepsilon|\alpha, X_1) F(d\alpha, dX_1) \\ &= \int g_0(x, \alpha, \varepsilon) F(d\varepsilon, d\alpha, dX_1) = \mu(x). \end{aligned}$$

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.

$$\int [m_0(\tilde{x}, \alpha) - m_0(\bar{x}, \alpha)] F(d\alpha) = \mu(\tilde{x}) - \mu(\bar{x}).$$

Through the rest of the paper we assume that the support of $X_i$ is finite (so $X_{it}$ is discrete). A useful example is binary $X_{it}$, where $X_{it} \in \{0, 1\}$. With discrete $X_{it}$ the 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

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

## 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

$$E[Y_{i,t_k}|X_i = X^k] = E[g_0(X_{it_k}, \alpha_i, \varepsilon_{it_k})|X_i = X^k] = E[g_0(x, \alpha_i, \varepsilon_{i1})|X_i = X^k],$$

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 $g_0(x, \alpha_i, \varepsilon_{it})$ 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 $B_\ell \leq g_0(x, \alpha_i, \varepsilon_{it}) \leq B_u$ for constants $B_\ell$ 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

$$E[Y_{i,t_k}|X_i = X^k] = E[g_0(X^k_{t_k}, \alpha_i, \varepsilon_{it_k})|X_i = X^k] = E[g_0(x, \alpha_i, \varepsilon_{i1})|X_i = X^k].$$

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

$$B_\ell \leq E[g_0(x, \alpha_i, \varepsilon_{i1})|X_i = X^k] \leq B_u.$$

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, as shown in Section 6.

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 show 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 $k \in \mathcal{K}(x)$,

$$E[1(g_0(x,\alpha_i,\varepsilon_{i1}) \leq y)|X_i = X^k] = E[1(Y_{i,t_k} \leq y)|X_i = X^k].$$

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

$$E[1(Y_{i,t_k} \leq y)|X_i = X^k] = E[1(g_0(X_t^k,\alpha_i,\varepsilon_{it_k}) \leq y)|X_i = X^k] = E[1(g_0(x,\alpha_i,\varepsilon_{i1}) \leq y)|X_i = X^k].$$

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

$$0 \leq E[1(g_0(x, \alpha_i, \varepsilon_{i1}) \leq y)|X_i = X^k] \leq 1.$$

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

$$\begin{aligned} G_\ell(y, x) &= \sum_{k \in \mathcal{K}(x)} \mathcal{P}^k E[1(Y_{i,t_k} \leq y)|X_i = X^k] = \sum_{k \in \mathcal{K}(x)} \mathcal{P}^k E[1(g_0(x, \alpha_i, \varepsilon_{i1}) \leq y)|X_i = X^k] \\ &\leq \sum_k \mathcal{P}^k E[1(g_0(x, \alpha_i, \varepsilon_{i1}) \leq y)|X_i = X^k] = G(y, x) \leq G_\ell(y, x) + \bar{\mathcal{P}}(x) = G_u(y, x). \end{aligned}$$

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

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

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

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 $X_i = (X_{i1}, ..., X_{iT})'$. Confidence intervals for the identified set can then be formed 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.

## 4 Time Effects in Static Models

In static models it is possible to relax the time homogeneity of $g_0(x, \alpha, \varepsilon)$ to allow for additive location and multiplicative scale time effects. These effects can even be allowed to depend on $x$, though we focus here on the case where they do not.

Consider a model where

$$Y_{it} = g_{t0}(X_{it}, \alpha_i, \varepsilon_{it}),\ g_{t0}(x, \alpha, \varepsilon) = \tau_t + \sigma_t g_0(x, \alpha, \varepsilon),\ \tau_1 = 0,\ \sigma_1 = 1, \tag{4}$$

and $\tau_t$ and $\sigma_t$ are period specific location and scale effects. We impose the restriction that the location effect is zero and the scale effect is one in the first time period. We continue to maintain the time homogeneity assumption of equation (2). Now the ASF and QSF depend on $t$ and are given by

$$\begin{aligned} \mu_t(x) &= \tau_t + \sigma_t \int g_0(x, \alpha, \varepsilon) F(d\varepsilon, d\alpha), \qquad (5) \\ q_t(\lambda, x) &= \lambda^{th} \text{ quantile of } \tau_t + \sigma_t g_0(x, \alpha_i, \varepsilon_{it}) \\ &= \tau_t + \sigma_t \cdot \lambda^{th} \text{ quantile of } g_0(x, \alpha_i, \varepsilon_{it}). \end{aligned}$$

We use the fact that $E[g_0(x,\alpha_i,\varepsilon_{it})|X_i]$ does not depend on $t$ to identify location and scale effects. Different time periods with the same $x$ provide identifying information for time effects. In particular, if $X_t^k = x$ and $X_1^k = x$ then

$$\begin{aligned} E[Y_{it}|X_i &= X^k] = \tau_t + \sigma_t E[g_0(x,\alpha_i,\varepsilon_{it})|X_i = X] \\ &= \tau_t + \sigma_t E[g_0(x,\alpha_i,\varepsilon_{i1})|X_i = X^k] = \tau_t + \sigma_t E[Y_{i1}|X_i = X^k]. \end{aligned}$$

Using two different $X^k$, or sets of $X^k$, then leads to identification of $\tau_t$ and $\sigma_t$. For example, consider the $T = 2$ model and binary $x$, where $x \in \{0,1\}$. Then for $X^k \in \{(0,0)', (1,1)'\}$,

$$\begin{aligned} E[Y_{i2}|X_i = (0,0)] = \tau_2 + \sigma_2 E[Y_{i1}|X_i = (0,0)], \\ E[Y_{i2}|X_i = (1,1)] = \tau_2 + \sigma_2 E[Y_{i1}|X_i = (1,1)]. \end{aligned}$$

This two equation system can be solved for the two unknowns $\tau_2$ and $\sigma_2$ as long as $E[Y_{i1}|X_i = (1,1)] \neq E[Y_{i1}|X_i = (0,0)]$.

In general, let

$$\bar{\mathcal{X}}_t = \{X : X_1 = X_t\}; t = 2, ..., T,$$

and partition $\bar{\mathcal{X}}_t$ into two disjoint sets $\bar{\mathcal{X}}_t^1$ and $\bar{\mathcal{X}}_t^2$. Then, similar to the previous example,

$$\begin{aligned} E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^1] = \tau_t + \sigma_t E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^1], \\ E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^2] = \tau_t + \sigma_t E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^2]. \end{aligned}$$

The location and scale effects are identified by solving these two equations for each $t$.

Theorem 4: *If equations (2) and (4) are satisfied,* $E[|Y_{it}|] < \infty$ *for all* $t$*, and* $\Pr(X_i \in \bar{\mathcal{X}}_t^j) > 0$ *and* $E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^1] \neq E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^2]$*, for each* $t = 2, ..., T; j = 1, 2$*, then*

$$\sigma_t = \frac{E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^2] - E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^1]}{E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^2] - E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^1]}, \tau_t = E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^2] - \sigma_t E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^2].$$

This result gives a very simple way to identify the time effects. In general, there may be multiple partitions $\bar{\mathcal{X}}_t^1$ and $\bar{\mathcal{X}}_t^2$ that work. In that case $\tau_t$ and $\sigma_t$ may be overidentified. For efficiency it would be desireable to estimate using optimal minimum distance. However, the small sample properties of this are likely to be poor because some data cells may have few observations, and so we focus on the simple partition into two sets.

The time varying ASF and QSF can be recovered by removing the identified location and scale effects in the bounds and then adding them back at each time period.

Theorem 5: *If equations (2) and (4) are satisfied,* $E[|Y_{it}|] < \infty$ *for all* $t$*,* $\Pr(X_i \in \bar{\mathcal{X}}_t^j) > 0$ *and* $E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^1] \neq E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^2]$ *for each* $t = 2, ..., T; j = 1, 2$*, and* $B_\ell \leq g_0(x,\alpha_i,\varepsilon_{it}) \leq B_u$ *for constants* $B_\ell$ *and* $B_u$ *and all* $x$*, then*

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

*where*

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

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

$$E[\frac{Y_{i,t_k}-\tau_{t_k}}{\sigma_{t_k}}|X_i = X^k] = E[g_0(X^k_{t_k}, \alpha_i, \varepsilon_{it_k})|X_i = X^k] = E[g_0(x, \alpha_i, \varepsilon_{i1})|X_i = X^k].$$

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

$$B_\ell \le E[g_0(x, \alpha_i, \varepsilon_{i1})|X_i = X^k] \le B_u.$$

The conclusion then follows by multiplying by $\mathcal{P}^k$, adding over $k$, multiplying by $\sigma_t$, and adding $\tau_t$. Q.E.D.

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

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

THEOREM 6: *If equations, (2), and (4) are satisfied,* $\Pr(X_i \in \bar{\mathcal{X}}^j_t) > 0$ *and* $E[Y_{i1}|X_i \in \bar{\mathcal{X}}^1_t] \neq E[Y_{i1}|X_i \in \bar{\mathcal{X}}^2_t]$ *for each* $t = 2, ..., T; j = 1, 2, T$, *then*

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

*where*

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

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

$$\begin{aligned}
E[1(\frac{Y_{i,t_k}-\tau_{t_k}}{\sigma_{t_k}} &\le y)|X_i = X^k] = E[1(g_0(X^k_{t_k}, \alpha_i, \varepsilon_{it_k}) \le y)|X_i = X^k] \\
&= E[1(g_0(x, \alpha_i, \varepsilon_{i1}) \le y)|X_i = X^k].
\end{aligned}$$

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

$$0 \le E[1(g_0(x, \alpha_i, \varepsilon_{i1}) \le y)|X_i = X^k] \le 1.$$

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

$$\begin{aligned}
G_\ell(y,x) &= \sum_{k\in\mathcal{K}(x)} \mathcal{P}^k E[1(\frac{Y_{i,t_k}-\tau_{t_k}}{\sigma_{t_k}} \le y)|X_i = X^k] \\
&= \sum_{k\in\mathcal{K}(x)} \mathcal{P}^k E[1(g_0(x,\alpha_i,\varepsilon_{i1}) \le y)|X_i = X^k] \\
&\le G(y,x) \le G_u(y,x) + \bar{\mathcal{P}}(x).
\end{aligned}$$

The conclusion then follows by inverting, multiplying by $\sigma_t$, and adding $\tau_t$. Q.E.D.

The QSF bounds are unusual in that the quantile time effects are identified from expectations. This approach depends crucially on $\tau_t$ and $\sigma_t$ being constant (i.e. nonrandom). The ASF bounds will also apply when $\tau_t$ and $\sigma_t$ are random and independent of the data, but the QSF bounds will not.

We can generalize this to the case where $\tau_t$ and $\sigma_t$ may depend on $x$. In this case the model is

$$Y_{it} = g_{t0}(X_{it},\alpha_i,\varepsilon_{it}),\ g_{t0}(x,\alpha,\varepsilon) = \tau_t(x) + \sigma_t(x) g_0(x,\alpha,\varepsilon),\ \tau_1(x)=0,\ \sigma_1(x)=1. \qquad (6)$$

The objects of interest will be the same as in equation (5), with an $x$ argument included for $\tau_t(x)$ and $\sigma_t(x)$. Let

$$\bar{\mathcal{X}}_t(x) = \{X : X_1 = X_t = x\}; t = 2,...,T,$$

and partition $\bar{\mathcal{X}}_t(x)$ into two disjoint sets $\bar{\mathcal{X}}_t^1(x)$ and $\bar{\mathcal{X}}_t^2(x)$. Such a partition may be possible except when $T=2$. Then similarly to the constant case,

$$\begin{aligned}
\sigma_t(x) &= \frac{E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^2(x)] - E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^1(x)]}{E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^2(x)] - E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^1(x)]}, \\
\tau_t(x) &= E[Y_{it}|X_i \in \bar{\mathcal{X}}_t^2(x)] - \sigma_t(x) E[Y_{i1}|X_i \in \bar{\mathcal{X}}_t^2(x)].
\end{aligned}$$

Thus $\tau_t(x)$ and $\sigma_t(x)$ are identified as in the conclusion of Theorem 4. Also, corresponding results to Theorem 5 and 6 follow, with $\tau_t(x)$ and $\sigma_t(x)$ replacing $\tau_t$ and $\sigma_t$ respectively.

Estimation of $\tau_t(x)$ and $\sigma_t(x)$ depending on $x$ will require that many data cells (corresponding to different values of $X_i$) have positive probability when $x$ takes on several values and $T$ is large. Practically speaking, it may be hard to identify such effects in data typically encountered in economics. For this reason we have focused on constant time effects here. Graham and Powell (2008) did consider a linear random coefficients model with location effects that depend on $x$.

Even if $\tau_t(x)$ and $\sigma_t(x)$ are allowed to depend on $x$ there may be overidentifying restrictions implied by the model with strict exogeneity. A characterization of all these restrictions is left to future research.

## 5 Bounds in the Dynamic Model

In static models we developed bounds by conditioning on the entire $X_i$ vector. The dynamic model only imposes independence from lagged $X_{it}$, so we will condition only on lagged $X_{it}$. Specifically, we partition of the support of $X_i$ into sets where the first occurrence of $x$ is at time $t$ and the set where $x$ never occurs. This partition is given by

$$\mathcal{X}_t(x) = \{X : X_t = x,\ X_s \neq x\ \forall s < t\}, t = 1, ..., T; \bar{\mathcal{X}}(x) = \{X : X_t \neq x\ \forall t\}.$$

There is a fundamental result that provides partial identification using this partition. Define

$$\mathcal{A}_t = \{X_i \in \mathcal{X}_t(x)).$$

Note that $\mathcal{A}_t$ only restricts $X_{it}, ..., X_{i1}$. Let $1(\mathcal{A}_t)$ be the indicator function for $\mathcal{A}_t$. For all $t$,

$$\begin{aligned} E[1(\mathcal{A}_t)g_0(x, \alpha_i, \varepsilon_{iT})] &= E[1(\mathcal{A}_t)E[g_0(x, \alpha_i, \varepsilon_{iT})|X_{iT}, ..., X_{i1}, \alpha_i]] \\ &= E[1(\mathcal{A}_t)E[g_0(x, \alpha_i, \varepsilon_{it})|X_{it}, ..., X_{i1}, \alpha_i]] \\ &= E[1(\mathcal{A}_t)g_0(x, \alpha_i, \varepsilon_{it})] = E[1(\mathcal{A}_t)Y_{it}] \end{aligned} \tag{7}$$

where the second equality follows by equation (3) and the last equality by $X_{it} = x$ for all $X_i \in \mathcal{A}_t$. Combining this result with the fact that the distribution of $(\alpha_i, \varepsilon_{it})$ does not vary with $t$ (also implied by equation (3)) leads to the following bounds:

Theorem 7: *If equations (1) and (3) are satisfied and for all $x$, and $B_\ell \leq g_0(x, \alpha_i, \varepsilon_{it}) \leq B_u$ for constants $B_\ell$ and $B_u$ and all $x$, then*

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

*where*

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

Proof of Theorem 7: Equation (3) implies that $\varepsilon_{it}$ is independent of $X_{it}, ..., X_{i2}$ conditional on $(\alpha_i, X_{i1})$, since the conditional distribution of $\varepsilon_{it}$ given $(X_{it}, ..., X_{i1}, \alpha_i)$ does not depend on $X_{it}, ..., X_{i2}$. It also implies that this distribution is the same for all $t$, being equal to that for $t = 1$. It follows that the distribution of $(\alpha_i, \varepsilon_{it})$ does not vary with $t$. Also by the sets being a partition we have $1(X_i \in \bar{\mathcal{X}}(x)) + \sum_t 1(\mathcal{A}_t) = 1$. Therefore, by eq. (7)

$$\begin{aligned} \mu(x) &= E[g_0(x, \alpha_i, \varepsilon_{iT})] = \sum_{t=1}^{T} E[1(\mathcal{A}_t)g_0(x, \alpha_i, \varepsilon_{iT})] + E[1(X_i \in \bar{\mathcal{X}}(x))g_0(x, \alpha_i, \varepsilon_{iT})] \\ &= \sum_{t=1}^{T} E[1(\mathcal{A}_t)Y_{it}] + E[1(X_i \in \bar{\mathcal{X}}(x))g_0(x, \alpha_i, \varepsilon_{iT})]. \end{aligned}$$

We also have

$$B_\ell \bar{\mathcal{P}}(x) \leq E[1(X_i \in \bar{\mathcal{X}}(x))g_0(x, \alpha_i, \varepsilon_{iT})] \leq B_u \bar{\mathcal{P}}(x).$$

Replacing $E[1(X_i \in \bar{\mathcal{X}}(x))g_0(x, \alpha_i, \varepsilon_{iT})]$ in these inequalities with $\mu(x) - \sum_{t=1} E[1(\mathcal{A}_t)Y_{it}]$ from the previous equation and adding $\sum_{t=1} E[1(\mathcal{A}_t)Y_{it}]$ to both inequalities gives the bounds. 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

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

The width of the bounds is $\Pr(X_i = (1, ..., 1)') + \Pr(X_i = (0, ..., 0)')$, that will tend to be large in short panels but more informative in long ones. This is a bounds solution to the problem of identifying state dependence in the presence of unobserved heterogeneity (Feller, 1943, and Heckman, 1981). Note that

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

is the effect of lagged $Y_{it}$, holding $\alpha_i$ fixed, averaged over $\alpha_i$. The conditional distribution of $Y_{it}$ is completely characterized by the two random variables $\Pr(Y_{it} = 1 | Y_{i,t-1} = 1, \alpha)$ and $\Pr(Y_{it} = 1 | Y_{i,t-1} = 0, \alpha)$, so that we can think of $\alpha_i$ as two dimensional, being equal to these two random variables. Our results put no restrictions on the joint distribution of these conditional probabilities.

For bounds for the QSF define

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

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

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

*where*

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

Proof of Theorem 8: Replacing $g_0(x, \alpha_i, \varepsilon_{iT})$ by $1(g_0(x, \alpha_i, \varepsilon_{iT}) \leq y)$ in eq. (7) gives

$$E[1(\mathcal{A}_t)1(g_0(x, \alpha_i, \varepsilon_{iT}) \leq y)] = E[1(\mathcal{A}_t)1(Y_{it} \leq y)].$$

Then proceding as in the proof of Theorem 7,

$$\begin{aligned} G(y, x) &= E[1(g_0(x, \alpha_i, \varepsilon_{iT}) \leq y)] \\ &= \sum_{t=1} E[1(\mathcal{A}_t)1(Y_{it} \leq y)] + E[1(X_i \in \bar{\mathcal{X}}(x))1(g_0(x, \alpha_i, \varepsilon_{iT}) \leq y)]. \end{aligned}$$

We also have

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

implying the bounds on $G(y, x)$. Inverting those bounds, e.g. similarly to Imbens and Newey (2009), gives the result. *Q.E.D.*

The bounds for the dynamic model also apply to the static model but there are advantages to using the static bounds when they apply. One advantage is that the bounds for the static model use more time periods, which should help reduce sampling variability in estimators.

# 6 Bounds under Monotonicity

When properties of $g_0$ are known it should be possible to tighten the bounds. We consider here the case of monotonicity, where it is known that for some $\tilde{x}$ and $\bar{x}$,

$$g_0(\tilde{x}, \alpha_i, \varepsilon_{it}) \geq g_0(\bar{x}, \alpha_i, \varepsilon_{it}). \tag{8}$$

This condition leads to tighter bounds for the ASF, QSF, and for treatment effects in the static and dynamic cases. To describe the bounds, recall that $\bar{\mathcal{K}}(x) = \{k : X_t^k \neq x \forall t\}$, and let $\bar{\mathcal{P}}(\bar{x}, \tilde{x}) = \Pr(\bar{\mathcal{K}}(\tilde{x}) \cap \bar{\mathcal{K}}(\bar{x}))$. For $k \in \mathcal{K}(\tilde{x}) \cap \mathcal{K}(\bar{x})$, let $\tilde{t}_k$ and $\bar{t}_k$ be time periods such that $X_{\tilde{t}_k}^k = \tilde{x}$ and $X_{\bar{t}_k} = \bar{x}$.

Theorem 9: *Suppose that $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ for $x \in \{\tilde{x}, \bar{x}\}$ and equations (1), (2), and (8) are satisfied. Let $\mathcal{A}_k = \{X_i = X^k\}$. Then*

$$\mu(\tilde{x}) - \mu(\bar{x}) \geq \sum_{k \in \mathcal{K}(\tilde{x}) \cap \mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)\{Y_{i\tilde{t}_k} - Y_{i\bar{t}_k}\}].$$

*If* $g_0(\tilde{x}, \alpha_i, \varepsilon_{it}) \geq B_\ell$ *then*

$$\mu(\tilde{x}) \geq \sum_{k \in \mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)Y_{i\tilde{t}_k}] + \sum_{k \in \bar{\mathcal{K}}(\tilde{x}) \cap \mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)Y_{i\bar{t}_k}] + \bar{\mathcal{P}}(\bar{x}, \tilde{x})B_\ell.$$

*If* $g_0(\bar{x},\alpha_i,\varepsilon_{it}) \leq B_u$ *then*

$$\mu(\bar{x}) \leq \sum_{k\in\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)Y_{i\bar{t}_k}] + \sum_{k\in\bar{\mathcal{K}}(\bar{x})\cap\mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)Y_{i\tilde{t}_k}] + \bar{\mathcal{P}}(\bar{x},\tilde{x})B_u.$$

Proof of Theorem 9: By the monotonicity condition,

$$\begin{aligned}
\mu(\tilde{x}) &= \sum_k E[1(\mathcal{A}_k)g_0(\tilde{x},\alpha_i,\varepsilon_{it})] \\
&\geq \sum_{k\in\mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)Y_{i\tilde{t}_k}] + \sum_{k\in\bar{\mathcal{K}}(\tilde{x})\cap\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)Y_{i\bar{t}_k}] + \sum_{k\in\bar{\mathcal{K}}(\tilde{x})\cap\bar{\mathcal{K}}(\bar{x})} E[1(\mathcal{A}_k)g_0(\tilde{x},\alpha_i,\varepsilon_{it})] \\
&\geq \sum_{k\in\mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)Y_{i\tilde{t}_k}] + \sum_{k\in\bar{\mathcal{K}}(\tilde{x})\cap\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)Y_{i\bar{t}_k}] + \bar{\mathcal{P}}(\bar{x},\tilde{x})B_\ell.
\end{aligned}$$

The last inequality gives the second conclusion. We also have

$$\begin{aligned}
\mu(\bar{x}) &= \sum_k E[1(\mathcal{A}_k)g_0(\bar{x},\alpha_i,\varepsilon_{it})] \\
&\leq \sum_{k\in\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)Y_{i\bar{t}_k}] + \sum_{k\in\bar{\mathcal{K}}(\bar{x})\cap\mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)Y_{i\tilde{t}_k}] + \sum_{k\in\bar{\mathcal{K}}(\tilde{x})\cap\bar{\mathcal{K}}(\bar{x})} E[1(\mathcal{A}_k)g_0(\bar{x},\alpha_i,\varepsilon_{it})] \\
&\leq \sum_{k\in\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)Y_{i\bar{t}_k}] + \sum_{k\in\bar{\mathcal{K}}(\bar{x})\cap\mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)Y_{i\tilde{t}_k}] + \bar{\mathcal{P}}(\bar{x},\tilde{x})B_u.
\end{aligned}$$

The last inequality gives the second conclusion. To obtain the first conclusion, subtract the second inequality here from the previous second inequality, to obtain

$$\begin{aligned}
\mu(\tilde{x}) - \mu(\bar{x}) &\geq \sum_{k\in\mathcal{K}(\tilde{x})\cap\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)\{Y_{i\tilde{t}_k} - Y_{i\bar{t}_k}\}] \\
&\quad + \sum_{k\in\bar{\mathcal{K}}(\tilde{x})\cap\bar{\mathcal{K}}(\bar{x})} E[1(\mathcal{A}_k)\{g_0(\tilde{x},\alpha_i,\varepsilon_{it}) - g_0(\bar{x},\alpha_i,\varepsilon_{it})\}] \\
&\geq \sum_{k\in\mathcal{K}(\tilde{x})\cap\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)\{Y_{i\tilde{t}_k} - Y_{i\bar{t}_k}\}]. Q.E.D.
\end{aligned}$$

A symmetric argument for the case $g_0(\bar{x},\alpha_i,\epsilon_{it}) \geq g_0(\tilde{x},\alpha_i,\epsilon_{it})$ gives an upper bound:

$$\mu_u(\tilde{x}) - \mu_\ell(\bar{x}) = \sum_{k\in\mathcal{K}(\tilde{x})\cap\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)(Y_{i\tilde{t}_k} - Y_{i\bar{t}_k})].$$

These bounds are the same as the bounds for the average partial effect under monotonicity in Chernozhukov et al. (2007).

Consider now the QSF. Define

$$\begin{aligned}
G_u^*(y,\tilde{x}) &= \sum_{k\in\mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)1(Y_{i\tilde{t}_k} \leq y)] + \sum_{k\in\bar{\mathcal{K}}(\tilde{x})\cap\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)1(Y_{i\bar{t}_k} \leq y)] + \bar{\mathcal{P}}(\bar{x},\tilde{x}), \\
G_\ell^*(y,\bar{x}) &= \sum_{k\in\mathcal{K}(\bar{x})} E[1(\mathcal{A}_k)1(Y_{i\bar{t}_k} \leq y)] + \sum_{k\in\bar{\mathcal{K}}(\bar{x})\cap\mathcal{K}(\tilde{x})} E[1(\mathcal{A}_k)1(Y_{i\tilde{t}_k} \leq y)],
\end{aligned}$$

$$q^*_\ell(\tilde{x},\lambda) = \begin{cases} -\infty, \lambda \le \bar{\mathcal{P}}(\bar{x},\tilde{x}), \\ (G^*_u)^{-1}(\lambda,\tilde{x}), \bar{\mathcal{P}}(\bar{x},\tilde{x}) < \lambda. \end{cases}, q^*_u(\lambda,\bar{x}) = \begin{cases} (G^*_\ell)^{-1}(\lambda,x), 0 < \lambda < 1-\bar{\mathcal{P}}(\bar{x},\tilde{x}), \\ +\infty, \lambda \ge 1-\bar{\mathcal{P}}(\bar{x},\tilde{x}). \end{cases}.$$

THEOREM 10: *If equations(1), (2), and (8) are satisfied then*

$$q(\lambda,\tilde{x}) \ge q^*_\ell(\lambda,\tilde{x}), q(\lambda,\bar{x}) \le q^*_u(\lambda,\bar{x}).$$

Proof of Theorem 10: Note that monotonicity implies that

$$1(g(\tilde{x},\alpha_i,\varepsilon_{it}) \le y) \le 1(g(\bar{x},\alpha_i,\varepsilon_{it}) \le y).$$

It then follows similarly to the proof of Theorem 9 that

$$\begin{aligned} G(y,\tilde{x}) &= \sum_k E[1(\mathcal{A}_k)1(g(\tilde{x},\alpha_i,\varepsilon_{it}) \le y)] \le G^*_u(y,\tilde{x}), \\ G(y,\bar{x}) &= \sum_k E[1(\mathcal{A}_k)1(g(\bar{x},\alpha_i,\varepsilon_{it}) \le y)] \ge G^*_\ell(\bar{x},y). \end{aligned}$$

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

Turning now to the dynamic model, to sharpen the bounds for the monotonic case we use a different partition than in Section 5. Define $\mathcal{X}_t(\tilde{x},\bar{x}) = \mathcal{X}_t(\tilde{x}) \cup (\bar{\mathcal{X}}(\tilde{x}) \cap \mathcal{X}_t(\bar{x}))$. The partition we use here to derive a lower bound for $\mu(\tilde{x})$ is is

$$\{\mathcal{X}_1(\tilde{x},\bar{x}),...,\mathcal{X}_T(\tilde{x},\bar{x}),\bar{\mathcal{X}}(\tilde{x}) \cap \bar{\mathcal{X}}(\bar{x})\}.$$

The partition we use to derive an upper bound for $\mu(\bar{x})$ is the same with $\tilde{x}$ and $\bar{x}$ interchanged. They are finer partitions than the one given above. The fundamental identification result of Section 5 and monotonicity imply that

$$\begin{aligned} E[1(X_i &\in \bar{\mathcal{X}}_t(\tilde{x},\bar{x}))g_0(\tilde{x},\alpha_i,\varepsilon_{iT})] \\ &= E[1(X_i \in \mathcal{X}_t(\tilde{x}))g_0(\tilde{x},\alpha_i,\varepsilon_{iT})] + E[1(X_i \in \bar{\mathcal{X}}(\tilde{x}) \cap \mathcal{X}_t(\bar{x}))g_0(\tilde{x},\alpha_i,\varepsilon_{iT})] \\ &\ge E[1(X_i \in \mathcal{X}_t(\tilde{x}))Y_{it}] + E[1(X_i \in \bar{\mathcal{X}}(\tilde{x}) \cap \mathcal{X}_t(\bar{x}))g_0(\bar{x},\alpha_i,\varepsilon_{iT})] \\ &= E[1(X_i \in \bar{\mathcal{X}}_t(\tilde{x},\bar{x}))Y_{it}]. \end{aligned}$$

This inequality leads to a sharper lower bound for $\mu(\tilde{x})$ and one that interchanges $\tilde{x}$ and $\bar{x}$ leads to an upper bound for $\mu(\bar{x})$.

THEOREM 11: *Suppose that $E[|g_0(x,\alpha_i,\varepsilon_{it})|] < \infty$ for $x \in \{\tilde{x},\bar{x}\}$ and equations (1), (3), and (8) are satisfied. Then*

$$\mu(\tilde{x}) - \mu(\bar{x}) \ge \sum_{t=1}^T E[\{1(X_i \in \bar{\mathcal{X}}_t(\tilde{x},\bar{x})) - 1(X_i \in \bar{\mathcal{X}}_t(\bar{x},\tilde{x}))\}Y_{it}].$$

*If* $g_0(\tilde{x},\alpha_i,\varepsilon_{it}) \geq B_\ell$ *then*

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

*If* $g_0(\bar{x},\alpha_i,\varepsilon_{it}) \leq B_u$ *then*

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

Proof of Theorem 11: By the equation preceding Theorem 11 and monotonicity we have

$$\begin{aligned} E[g_0(\tilde{x},\alpha_i,\varepsilon_{iT})] &= \sum_{t=1}^{T} E[1(X_i \in \bar{\mathcal{X}}_t(\tilde{x},\bar{x}))g_0(\tilde{x},\alpha_i,\varepsilon_{iT})] + E[1(X_i \in \bar{\mathcal{X}}(\tilde{x}) \cap \bar{\mathcal{X}}(\bar{x}))g_0(\tilde{x},\alpha_i,\varepsilon_{iT})] \\ &\geq \sum_{t=1}^{T} E[1(X_i \in \bar{\mathcal{X}}_t(\tilde{x},\bar{x}))Y_{it}] + E[1(X_i \in \bar{\mathcal{X}}(\tilde{x}) \cap \bar{\mathcal{X}}(\bar{x}))g_0(\bar{x},\alpha_i,\varepsilon_{iT})]. \end{aligned}$$

By the analogous equation with $\bar{x}$ and $\tilde{x}$ interchanged,

$$\begin{aligned} E[g_0(\bar{x},\alpha_i,\varepsilon_{iT})] &= \sum_{t=1}^{T} E[1(X_i \in \bar{\mathcal{X}}_t(\bar{x},\tilde{x}))g_0(\bar{x},\alpha_i,\varepsilon_{iT})] + E[1(X_i \in \bar{\mathcal{X}}(\tilde{x}) \cap \bar{\mathcal{X}}(\bar{x}))g_0(\bar{x},\alpha_i,\varepsilon_{iT})] \\ &\leq \sum_{t=1}^{T} E[1(X_i \in \bar{\mathcal{X}}_t(\bar{x},\tilde{x}))Y_{it}] + E[1(X_i \in \bar{\mathcal{X}}(\tilde{x}) \cap \bar{\mathcal{X}}(\bar{x}))g_0(\bar{x},\alpha_i,\varepsilon_{iT})]. \end{aligned}$$

Subtracting these two inequalities gives the first conclusion. The second and third conclusions then follow as in the proof of Theorem 9. Q.E.D.

For quantile bounds with monotonicity in the dynamic case, let

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

$$q_\ell^*(\tilde{x},\lambda) = \begin{cases} -\infty, \lambda \leq \bar{\mathcal{P}}(\bar{x},\tilde{x}), \\ (G_u^*)^{-1}(\lambda,\tilde{x}), \bar{\mathcal{P}}(\bar{x},\tilde{x}) < \lambda. \end{cases} .q_u^*(\lambda,\bar{x}) = \begin{cases} (G_\ell^*)^{-1}(\lambda,x), 0 < \lambda < 1 - \bar{\mathcal{P}}(\bar{x},\tilde{x}), \\ +\infty, \lambda \geq 1 - \bar{\mathcal{P}}(\bar{x},\tilde{x}). \end{cases} .$$

THEOREM 12: *If equations (1), (3), and (8) are satisfied then*

$$q(\lambda,\tilde{x}) \geq q_\ell^*(\lambda,\tilde{x}), q(\lambda,\bar{x}) \leq q_u^*(\lambda,\bar{x}).$$

Proof of Theorem 12: By monotonicity it follows similarly to the proof of Theorem 11 that

$$\begin{aligned} G(y,\tilde{x}) &= \sum_{t=1}^{T} E[1(X_i \in \bar{\mathcal{X}}_t(\tilde{x},\bar{x}))1(g_0(\tilde{x},\alpha_i,\varepsilon_{iT}) \leq y)] \\ +E[1(X_i &\in \bar{\mathcal{X}}(\tilde{x}) \cap \bar{\mathcal{X}}(\bar{x}))1(g_0(\tilde{x},\alpha_i,\varepsilon_{iT}) \leq y)] \\ &\leq G_u^*(y,\tilde{x}), G(y,\bar{x}) \geq G_\ell^*(\bar{x},y). \end{aligned}$$

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

If the regressor is binary, $X_{it} \in \{0,1\}$, $\tilde{x} = 1$ and $\bar{x} = 0$, then $\bar{\mathcal{P}}(\bar{x}, \tilde{x}) = 0$. When the regressor takes on more than two values we can get tighter bounds if a monotonicity restriction holds for every possible pair of values. For example, if $x$ were a scalar and $g_0(\tilde{x}, \alpha_i, \epsilon_{it}) \geq g_0(\bar{x}, \alpha_i, \epsilon_{it})$ for *every* $\tilde{x}$ and $\bar{x}$ with $\tilde{x} > x$ then we could obtain improved bounds on the ASF and QSF.

# 7 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 13: *Suppose that equations (1) and (3) are satisfied, $\overrightarrow{X}_i = (X_{i1}, X_{i2}, ...)$ is stationary and, conditional on $\alpha_i$, the support of each $X_{it}$ is the marginal support of $X_{it}$ and $\overrightarrow{X}_i$ is ergodic. If $B_\ell \leq g_0(x, \alpha_i, \varepsilon_{it}) \leq B_u$ for constants $B_\ell$ 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 13: 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 $\alpha_i$ 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 $\alpha_i$. 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 $1(g_0(x, \alpha_i, \varepsilon_{it}) \leq y)$ 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 $\alpha_i$. 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 $\Pr(X_{it} \neq x|\alpha_i) = 1$ with positive probability, that is $\Pr(X_{it} = x|\alpha_i) = 0$ with positive probability. 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 14: *Suppose that equations (1) and (3) are satisfied, $\overrightarrow{X_i}$ is stationary and Markov of order $J$ conditional on $\alpha_i$, and for some $\varepsilon > 0$,*

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

*Then if* $B_\ell \leq g_0(x, \alpha_i, \varepsilon_{it}) \leq B_u$,

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

*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 14: 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 13. 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 $\alpha_i$ has any distribution with a compact support, Theorem 14 implies that the bounds shrink exponentially fast in $T$.

# 8 An Empirical Example

In this section we revisit the empirical question of how unions impact the wage structure using panel data. Our major contribution here is to estimate the effect without imposing the assumption that unobserved heterogeneity is some additive term that can be simply differenced out. In our model unobserved heterogeneity can have an almost unrestricted impact on the structural/causal response functions, with the time homogeneity serving as the only restriction. In our view, this constitutes a major step forward in answering this empirical question.

The effect of unions on wage structure is a longstanding question in labor economics – see Freeman (1984), Lewis (1986), Robinson (1989), Green (1991), and Card, Lemieux, and Riddell (2004) for surveys and additional references. Most previous empirical studies recognize the presence of unobserved differences between union and nonunion workers. For instance, in an influential study, Chamberlain (1982) finds strong evidence of heterogeneity bias in the estimation of the union effect by comparing estimates of cross- section models and panel data models with additive heterogeneity. This finding demonstrates the important need of controlling for unobserved heterogeneity. On the other hand, Angrist and Newey (1991) reject the hypothesis that the unobserved heterogeneity acts solely in an additive fashion. Thus, this finding demonstrates the important need of controlling for unobserved heterogeneity acting non-additively. Our tools and our study address precisely both of these needs.

We use data from the National Longitudinal Survey (Youth Sample). The sample consists of full-time young working males, 20 to 29 year-old in 1986, followed over the period 1986 to 1993. We exclude individuals who failed to provide sufficient information for each year, were in the active army forces or students any year, or reported too high (more than \$500 per hour) or too low (less than \$1 per hour) wages. The final sample includes 2,065 men. We use the union membership and the log hourly wage rate in 1980 dollars as the covariate and the outcome variables. The union membership variable reflects whether or not the individual had his wage set in collective bargaining agreement. We report results for panels with 2, 4, 6, and 8 years, all starting in 1986.

In our analysis, we focus on estimating the union effect for the subpopulation of workers that became ever unionized within the sample. For this subpopulation, the union effect is not point-identified, since there are 13% of the workers that stayed always unionized between 1986 and 1993. However, we hope to construct informative bounds on the union effect. We consider both a static model that allows for the union membership decisions to be strictly exogenous with respect to wage setting decisions, and a dynamic model that allows for the union membership decisions to be only predetermined with respect to wage setting decisions. We shall also report the estimates of the union effect for the subpopulation of workers who change the union status at least once within the sample. For this subpopulation, the effect is point-identified, that is, the bounds on the union effect collapse to a point. Finally, we shall not estimate the union effect for the entire population of workers, since the bounds are completely uninformative in this case. This happens because a substantial fraction of workers never changes the union status within the sample (see Table 1).

We begin by presenting the estimates of the union effect for the subpopulation of workers who change the union status at least once within the sample. In Figure 1 we compare our panel data estimates of quantile effects with the pooled cross-section estimates. In the cross-section estimates, we see that the quantile effect of union is positive but declines sharply at the upper end of the distribution, which agrees with previous cross section findings (Chamberlain, 1994). A common explanation for this phenomenon is that the high-skill workers at the lower end of the earning distribution tend to join the union, whereas the high-skill workers at the high end of the earning distribution tend not to join the union. The estimated quantile effect in the cross-section therefore captures this selection effect of unobserved skills. In the panel data estimates, which control for the unobserved skills, we see that the quantile effects of union become very flat across the quantile indices. Thus, by controlling for individual heterogeneity, we have eliminated the selection effect. Finally, our estimates of quantile effects are higher in the dynamic model than in the static model indicating a possible dynamic feedback between the wage setting and union membership decisions.

We next present estimated bounds on the union effect for the subpopulation of workers that became ever unionized within the sample. In Figures 2 and 3 we show these bounds for both static and dynamic models and for panels of lengths $T \in \{2, 4, 6, 8\}$ . In both cases, the size of the bounds decreases substantially with $T$. The bounds for $T = 8$ are informative, and show that the effect is positive for most of the quantile indices. In Figures 2 and 3, we also show bounds obtained using the assumption of monotonic and positive union effect on earnings. These bounds are also informative, and in fact are substantially tighter than the bounds obtained without the assumption of monotonicity.

Figure 4 plots 90% uniform confidence bands for the identified union effect and quantile union effect on ever unionized workers in the static and dynamic models. They are constructed by bootstrap with 500 repetitions. These bands allow us to make visual simultaneous inference on the entire quantile functions. For example, we cannot reject that the identified union effect is constant and positive for all the quantiles. For the ever unionized, the quantile union effect is positive for a large range of quantiles. The bands are narrower in the static model because this model uses more observations in the estimation of the quantile functions.

# 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] Angrist, J. D., and W. K. Newey (1991) "Over-Identification Tests in Earnings Functions With Fixed Effects," *Journal of Business and Economic Statistics* 9, 317-323.

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

[4] Bester, C.A., and C. Hansen (2009), "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Economic and Business Statistics* 27, 235-250.

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

[6] 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.

[7] Card, D., T. Lemieux and W. C. Riddell (2004) "Unionization and Wage Inequality: A Comparative Study of the U.S., the U.K., and Canada," unpublished manuscript.

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

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

[10] Chamberlain, G. (1994): "Quantile Regression, Censoring, and the Structure of Wages," in C. A. Sims (ed.), *Advances in Econometrics, Sixth World Congress,* Volume 1. Cambridge University Press. Cambridge.

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

[12] 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.

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

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

[15] Freeman, R. B. (1984), "Longitudinal Analysis of the Effect of Trade Unions," *Journal of Labor Economics* 2, 1-26.

[16] Green, D. A.(1991), "A Comparison of Estimation Approaches for the Union-Nonunion Wage Differential," *unpublished manuscript*, The University of British Columbia.

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

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

[19] 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.

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

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

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

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

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

[25] Lewis, H. G. (1986), *Union Relative Wage Effects: A Survey.* University of Chicago Press.

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

[27] Robinson, C., (1989), "The Joint Determination of Union Status and Union Wage Effect: Some Tests of Alternative Models, " *Journal of Political Economy* 97, 639-667.

[28] 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.

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

**Table 1: Empirical probabilities of union sequences**

| | Full sample | | Ever unionized |
|---|---|---|---|
| | Never unionized | Always unionized | Always unionized |
| T = 2 | 0.69 | 0.13 | 0.42 |
| T = 4 | 0.61 | 0.08 | 0.22 |
| T = 6 | 0.56 | 0.07 | 0.16 |
| T = 8 | 0.53 | 0.06 | 0.13 |

Source: NLSY79 1986-1993, 2,065 men. All the panels start in 1986

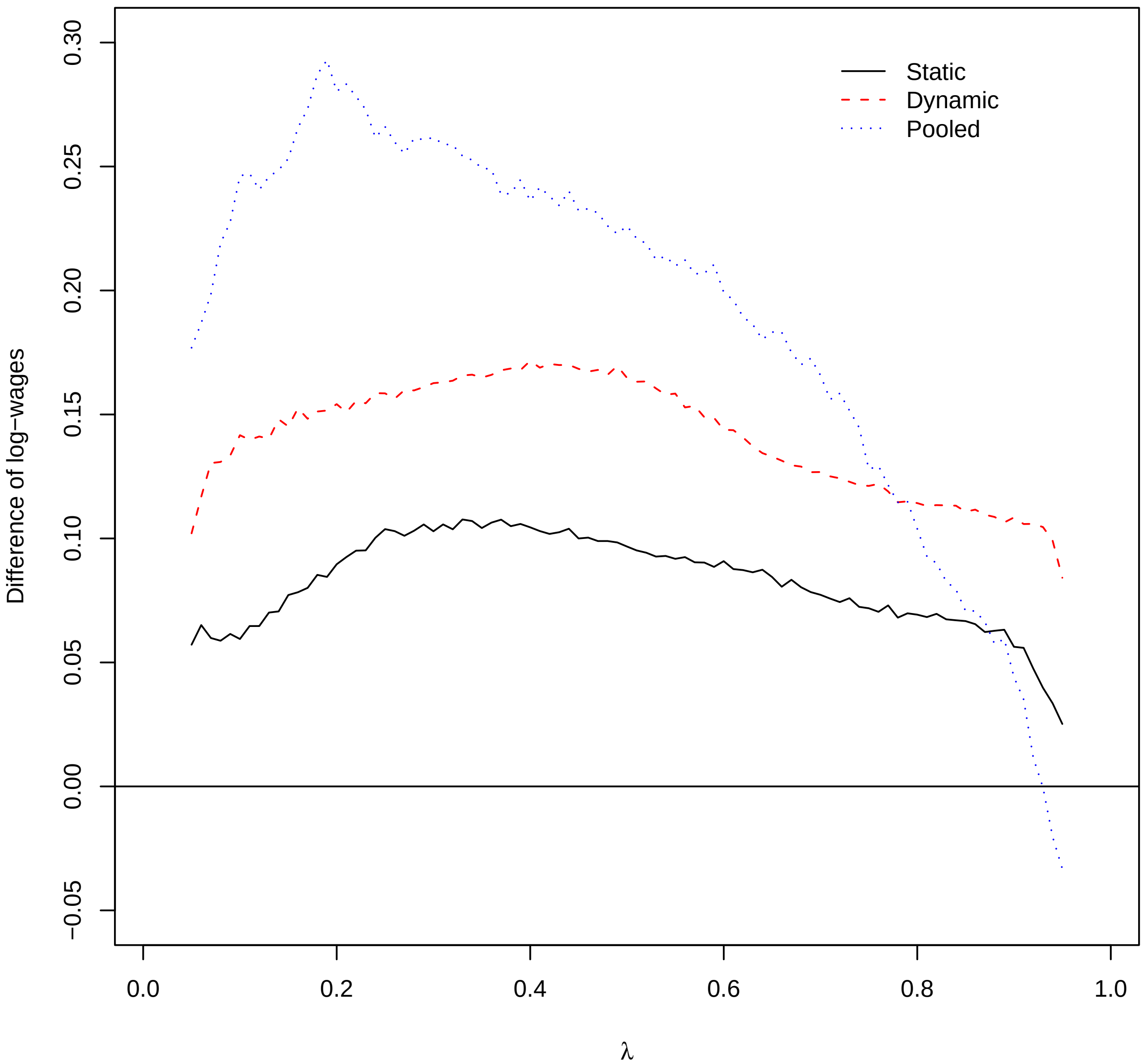


Figure 1: Identified quantile union effect. Estimates based on the entire panel 1986–1993.

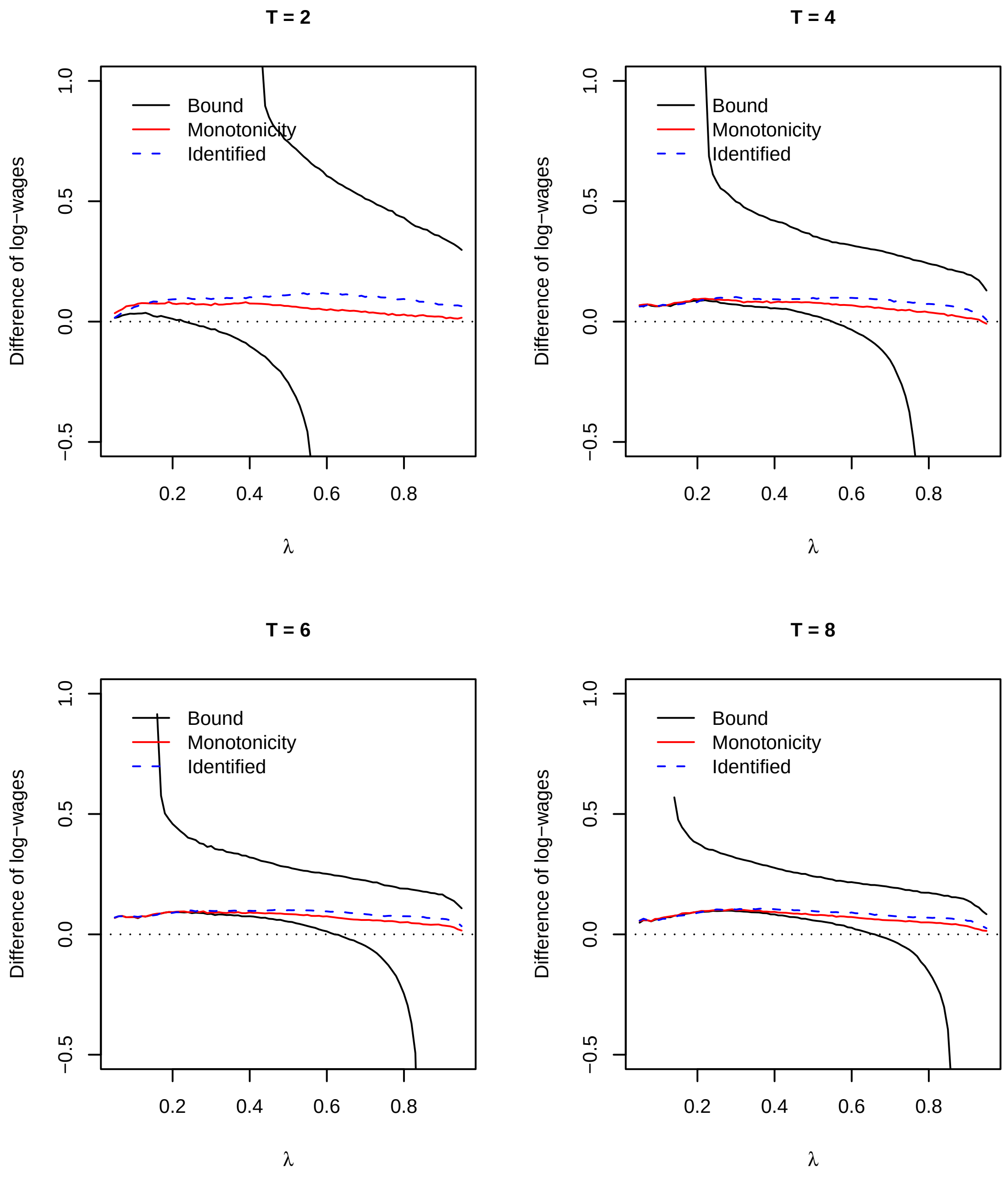


Figure 2: Bounds for quantile union effect on ever unionized. Static model.

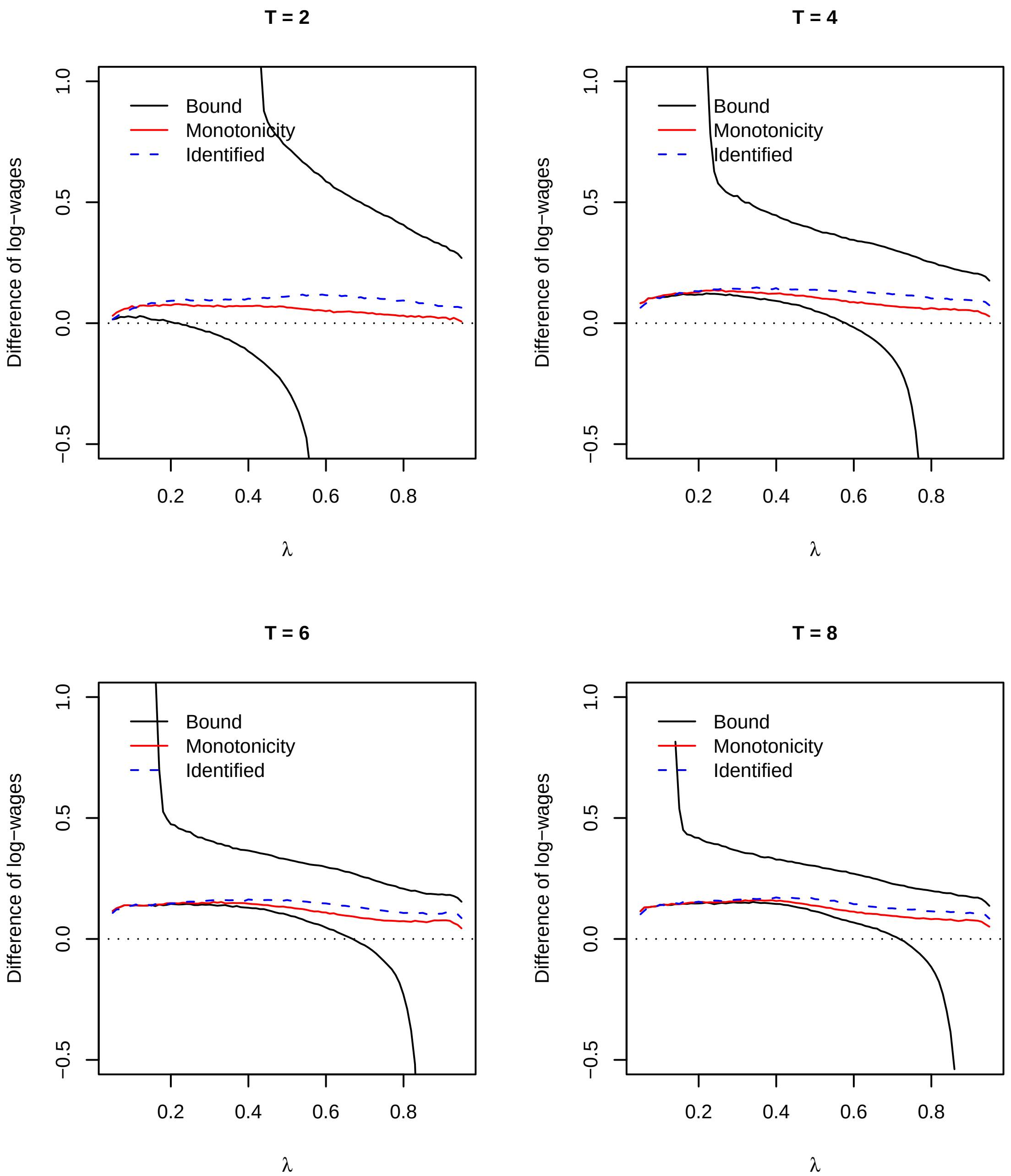


Figure 3: Bounds for quantile union effect on ever unionized. Dynamic model.

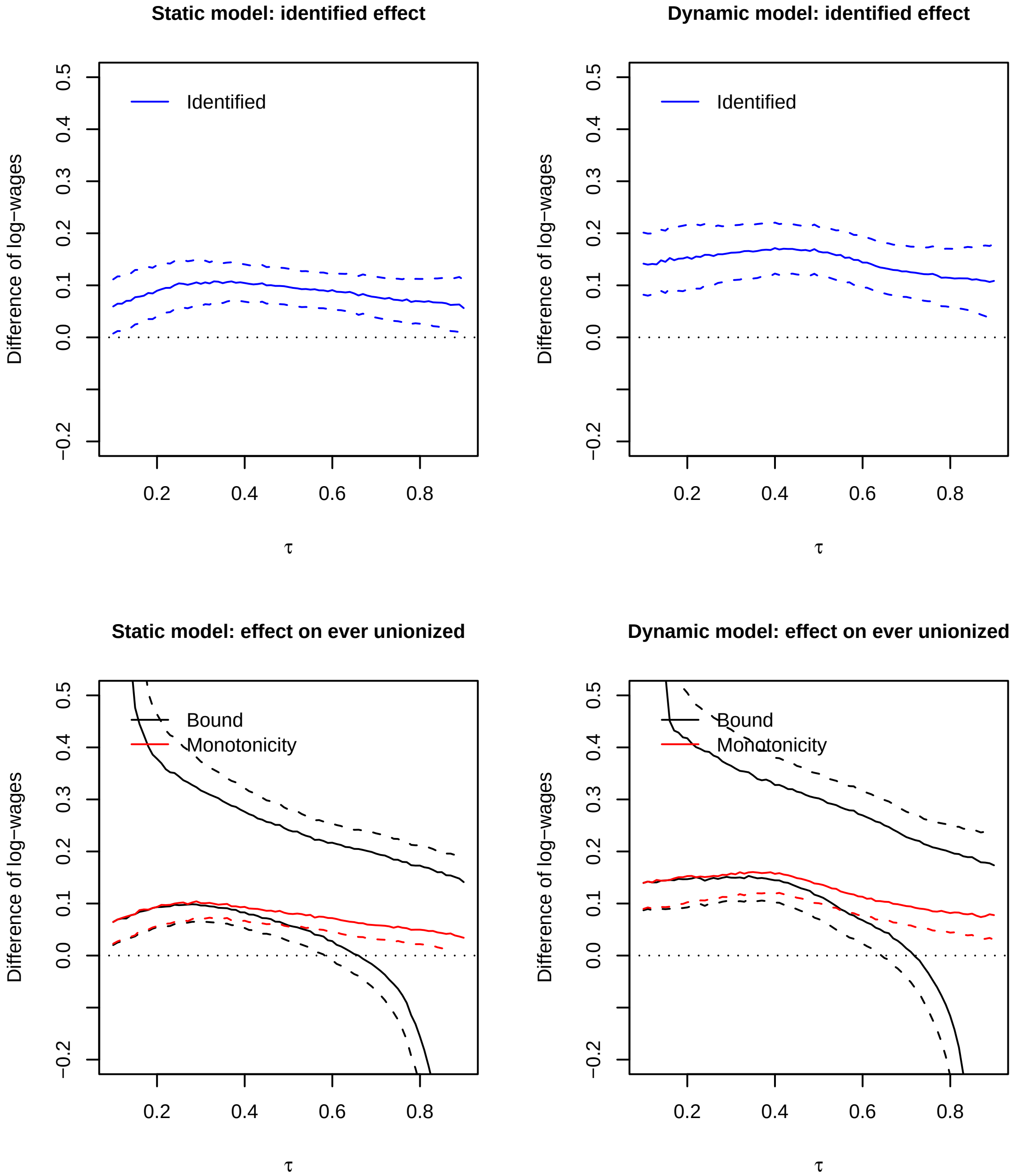


Figure 4: 90% bootstrap uniform confidence bands for the identified union effect and union effect on ever unionized (dashed lines). Estimates based on the entire panel 1986–1993.