Semiparametric identification and estimation of correlated random coefficient models for panel data[1]

Bryan S. Graham† and James L. Powell◇

Initial Draft: December 2007

This Draft: February 25, 2008

Preliminary and Incomplete

Comments Welcome

Abstract

In this paper we study identification and estimation of the causal effect of a small change in an endogenous regressor on a continuously-value outcome of interest using panel data. Specifically we focus on averages over the population distribution of unobserved heterogeneity, or average partial effects (APEs), and averages over subpopulations defined by their regressor values, or local average responses (LARs). Our central model assumes that the outcome variable is related to a (scalar) regressor, where the intercept and slope coefficients vary across individuals and are not independent of the regressor; for this model, we show how two measures of the outcome and regressor for each unit are sufficient for identification of the partial effects. This model is a semiparametric extension of the textbook linear fixed effects (FE) model widely used in empirical research; a distinctive feature of our approach is that it semiparametrically just identifies the APE and LAR and hence clearly illustrates the value and limits of panel data in dealing with endogeneity. A strengthening of our basic assumptions also allows us to identify Quantile Partial Effects (QPEs). We discuss extensions of our approach to models with multiple regressors and more than two time periods, and to models which permit contemporaneous endogeneity of the regressors and time-specific error terms.

JEL Classification: C14, C23, C33
Key Words: Panel Data, Correlated Random Coefficients, Control Function, Average Partial Effects, Local Average Response

[1] We would like to thank members of the Berkeley Econometrics Reading Group, Gary Chamberlain and Michael Jansson for helpful discussions. All the usual disclaimers apply.

† Department of Economics, University of California - Berkeley, 508-1 Evans Hall #3880, Berkeley, CA 94720 and National Bureau of Economic Research. E-mail: bgraham@econ.berkeley.edu. Web: http://www.econ.berkeley.edu/~bgraham/.

◇ Department of Economics, University of California - Berkeley, 508-1 Evans Hall #3880, Berkeley, CA 94720. E-mail: powell@econ.berkeley.edu. Web: http://www.econ.berkeley.edu/~powell/.

1 Introduction

That the availability of multiple observations of the same sampling unit (e.g., individual, firm, etc.) over time can help to control for the presence of unobserved heterogeneity is both intuitive and plausible. The inclusion of unit-specific intercepts in a linear regression model is widespread in empirical work (e.g., Griliches 1979, Currie and Thomas 1995, Card 1996, Burgess and Pande 2005). Researchers often combine such 'fixed effect' analyses with the method of two-stage least squares (e.g., Duflo 2001, Finkelstein 2002) or related producedures (e.g., Ashenfelter and Krueger 1994, Chay and Greenstone 2005).

The appropriateness of these modelling strategies hinges on any time-invariant correlated heterogeneity entering the outcome equation additively. Additivity, while statistically convenient, is difficult to motivate economically (cf., Card 2001, Imbens 2007).[2] Browning and Carro (2007) present a number of empirical panel data examples where non-additive forms of unobserved heterogeneity appear to be empirically relevant.

Recently Altonji and Matzkin (2005) have proposed methods for using panel data to control for nonseparable unobserved heterogeneity. Bester and Hansen (2007) develop related methods. These papers can be viewed as semiparametric generalizations of Chamberlain's (1980) correlated random effects (CRE) estimator (cf., Mundlak 1978, Newey 1994a). Identification follows from the imposition of semiparametric restrictions on the conditional distribution of the time-invariant heterogeneity given all leads and lags of the regressors.

In this paper we take a different, complementary, approach. As in Altonji and Matzkin (2005), our goal is to infer the effect of an exogenous change in the endogenous regressor on the outcome of interest. However, as in parametric fixed effects analyses, we leave the distribution of any time-invariant heterogeneity unrestricted. Instead, we secure identification by imposing restrictions on the form of the structural outcome equation and the distribution of time varying heterogeneity. The outcome equation is assumed to take a correlated random coefficients (CRC) form. Time-varying heterogeneity is assumed to satisfy a marginal stationarity restriction as in Manski (1987).

In the central model we consider, the availability of panel data is sufficient to deal with any endogeneity, though our approach can be extended to allow for additional 'triangular endogeneity' (cf., Blundell and Powell 2003, Imbens and Newey 2002). The estimators we propose for these two models can be viewed as semiparametric generalizations of the linear fixed effects (FE) and fixed effects instrumental variables (FE-IV) estimators (cf., Wooldridge 2005b, Murtazashvili and Wooldridge 2008).

To describe the models we consider, as well as our main results, more precisely let $Z_t = (Y_t, X_t')'$ be a random variable measured in each of $t = 1, \ldots, T$ periods for N randomly sampled units. The *structural outcome equation* is given by

$$Y_t = m\left(X_t, A, U_t\right) \tag{1}$$

[2] Chamberlain (1984) presents several well-formulated economic models that *do* imply linear specifications with unit-specific intercepts.

where Y_t is a scalar continuously-valued outcome of interest, X_t a choice variable, A time-invariant unobserved unit-level heterogeneity or 'fixed effects' and U_t a time-varying disturbance. Both A and U_t may be vector-valued. Since both time-invariant and time-varying heterogeneity enter (1) nonseparably the return to a marginal change in X_t is unit- and time-specific.

We are interested in estimands that characterize the average relationship between Y_t and X_t. Averaging may occur over the population distribution of unobserved heterogeneity, (A, U_t), or with respect to some well-defined subpopulation. For concreteness we focus on identification and estimation of the *average partial effect* (APE) (cf., Chamberlain 1984, Blundell and Powell 2003, Imbens and Newey 2002, Wooldridge 2005a) and the *local average response* (LAR) (cf., Altonji and Matzkin 2005, Bester and Hansen 2007). In the binary regressor case these two objects correspond to the average treatment effect (ATE) and the average treatment effect on the treated (ATT).

The average partial effect is given by

$$\gamma_t(x_t) = \mathbb{E}\left[\frac{\partial m(x_t, A, U_t)}{\partial x_t}\right], \tag{2}$$

where we assume that $m(x_t, a, u_t)$ is differentiable in its first argument. The local average response gives the average partial effect within a subpopulation defined by its choice; it is given by

$$\delta_t(x_t) = \mathbb{E}\left[\left.\frac{\partial m(X_t, A, U_t)}{\partial x_t}\right| X_t = x_t\right]. \tag{3}$$

Identification and estimation of (2) and (3) is nontrivial because X_t may vary systematically with A and/or U_t. The derivative of the regression function of Y_t given X_t does not identify $\delta(x_t)$. Differentiating through the integral we have

$$\frac{\partial E[Y_t|X_t = x_t]}{\partial x_t} = \delta_t(x_t) + \int\int m(x_t, a, u_t)\frac{\partial g(a, u_t|x_t)}{\partial x_t}\mathrm{d}m(a)\,\mathrm{d}m(u_t).$$

The second term is what Chamberlain (1982) calls heterogeneity bias.

Our central model imposes a CRC structure on (1) and marginal stationarity on the conditional distribution of U_t given A and $X = (X_1', \ldots, X_T')'$. When X_t is discretely-valued these assumptions generally only bound the APE and LAR (appropriately defined to account for the discreteness of X_t). Our analysis of this case also suggests new interpretations of the probability limits of the linear fixed effects (FE) estimator and the 'difference-in-differences' (DID) estimator of the program evaluation literature (Meyer 1995, Angrist and Krueger 1999, Athey and Imbens 2006). When X_t is continuously valued, and we additionally assume that the conditional distribution function of A given $X = x$ varies smoothly in x, the APE and LAR are point identified.

An extension of our central model also maintains a CRC structure on (1), but allows for the possibility of additional 'triangular' endogeneity. We augment (1) with a first-stage or *selection equation* relating the choice variable X_t to a vector of time-varying 'instruments', W_t:

$$X_t = g_t(W, B, V_t^*), \tag{4}$$

where B is a time-invariant scalar unobserved individual-specific intercept and V_t^* a mean-zero, time-varying, scalar disturbance. The vector of instruments may be discretely- or continuously-valued, however, for this model we assume that X_t is continuously-valued. We again study identification and estimation of (2) and (3). Our approach involves imposing restrictions on the conditional distribution of (U_t, V_t^*, B) given $W = (W_1', \ldots, W_T')'$ and A. In their working paper, Altonji and Matzkin (2001) study a similar system. Their approach involves restrictions on the distribution of (U_t, V_t^*, A) given W, while ours leaves, beyond smoothness assumptions, the distribution of A given W unrestricted.

In related work Wooldridge (2005b) and Murtazashvili and Wooldridge (2008) provide conditions under which the standard linear FE and FE-IV estimators are consistent for the APE. Under our assumptions these estimators are not consistent for the APE, but a straightforward strengthening of them recovers their results. Chernozhukov, Fernández-Val, Hahn and Newey (2007) study estimation of the ATE in a probit model with unit-specific intercepts (in the index function) and a binary regressor. They show that the maximum likelihood estimator which estimates the unit-specific intercepts along with the coefficient on X_t can be used to construct bounds on the ATE despite the incidental parameters problem (cf., Hahn 2001). Porter (1996) and Das (2003) study nonparametric estimation of panel data model with additive unobserved heterogeneity. Honoré (1992) and Abrevaya (2000) considers models with nonseparable heterogeneity but, like Manski (1987), only identify index coefficients, not the APE or LAR.

Our approach is distinctive from prior research in several ways. First, unlike Altonji and Matzkin (2005) and Bester and Hansen (2007), we treat A as a 'fixed effect.' Second, we show that our assumptions on the outcome equation and distribution of time-varying heterogeneity semiparametrically just-identify the APE and LAR. Put differently, the richest versions of our model, in contrast to those of Altonji and Matzkin (2005) and Bester and Hansen (2007), have no testable implications. Formally, we show that the maximum identification-preserving order of the CRC polynomial depends on time dimension of the panel. We view this feature of our approach positively, although acknowledge that others' may not. When the object of interest is the APE or the LAR, our approach clearly illustrates the value and limits of panel data for dealing with unobserved heterogeneity.[3] Researchers uncomfortable with the restrictions required for point identification can choose to construct bounds, change the estimands or undertake additional data collection (e.g., lengthen the T dimension of their panel). Finally, we develop methods which address the presence of residual triangular endogeneity. In their working paper Altonji and Matzkin (2001) also discuss a triangular system. Their model, in contrast in ours, does not explicitly allow for time-invariant correlated heterogeneity in the first-stage and also treats A as a correlated random effect.

Altonji and Matzkin (2005) and Bester and Hansen (2007), by imposing more restrictions on the conditional distribution of (A, U_t) given X are able to identify models with a nonparametric outcome equation, while our CRC outcome equation is semiparametric. For this reason we view our methods as complementary to theirs. We provide a stylized comparison of the two approaches

[3] Chesher (2007) also discusses the value of just-identifying conditions.

below.

In ongoing research we are extending our results to censored and/or discrete outcome variables. Unsurprisingly, and as is true with parametric nonlinear panel data models, simultaneously leaving the distribution of time-invariant heterogeneity unrestricted and identifying either the APE or LAR appears impossible.[4] Nevertheless marginal stationarity and a semiparametric model for the correlated random effects can be combined to achieve just identification of the APE and LAR. We illustrate some of the issues involved by means of an example in the conclusion.

The next section reports identification results for our central model, followed by a section that proposes consistent estimators for the APE under different distributional assumptions on the regressor. Section 4 considers extensions of our approach to estimation of quantile effects, the triangular model for endogenous regressors, and models with multiple regressors and time periods. Section 5 summarizes and suggests areas for further research.

2 Single regressor, two time periods

We illustrate each of our main results in the two period case and generalize them to panels of arbitrary length below. We assume that the structural outcome equation takes a correlated random coefficients form (CRC) form.

Assumption 2.1 (Correlated Random Coefficients)

$$Y_t = a\left(A, U_t\right) + b\left(A, U_t\right) X_t.$$

Given the CRC structure, the local average response of a change in X_t on Y_t is given by $\delta_t\left(x_t\right) = \mathbb{E}\left[b\left(A, U_t\right) \middle| X_t = x_t\right]$ and the average partial partial effect by $\gamma_t\left(x_t\right) = \mathbb{E}\left[b\left(A, U_t\right)\right]$. Note that, due to linearity at the unit-level, $\gamma_t\left(x_t\right)$ is constant in x_t.

Our key identifying assumption is marginal stationarity of the time-varying unobserved heterogeneity.

Assumption 2.2 (Marginal Stationarity) *(i)*

$$U_t | X, A \stackrel{D}{=} U_s | X, A, \qquad t \neq s,$$

(ii) the distribution of U_t *given* X *and* A *is non-degenerate for all* $(X, A) \in \mathcal{X} \times \mathcal{A}$.

Assumption 2.2 does not restrict the conditional distribution of A given X. In this sense A is a 'fixed effect'. Nevertheless Assumption 2.2, while allowing for serial dependence in U_t and certain forms of heteroscedasticity, is restrictive. For example it rules out heteroscedasticity over time (cf., Arellano 2003).

[4] Fixed effects estimators are available in certain parametric nonlinear models, but they do not, in general, allow the econometrician to infer the effect of an exogenous change in the endogenous regressor on the entire probability distribution of the outcome. Instead only certain features of this relationship are identified (e.g., ratios of the average partial effect of two regressors) (cf., Chamberlain 1984, Arellano and Honoré 2001, Arellano 2003).

To formally close the model we make the following sampling assumption:

Assumption 2.3 (Random Sampling) $\{(X_{1i}, X_{2i}, Y_{1i}, Y_{2i}, A_i)\}_{i=1}^{\infty}$ *is an independently and identically distributed random sequence drawn from the distribution* F_0.

Let $\beta_t(x) = \mathbb{E}[b(A, U_t)|X = x]$ denote the average effect of a small change in x_t within the subpopulation of units with $X = x = (x_1, x_2)'$. Observe that $\beta_t(x)$, while closely related, is the distinct from the LAR. It gives the average effect within a subpopulation defined by its entire *history* of choices for X_t. Our first result shows that $\beta_t(x)$ is just-identified when $x_1 \neq x_2$.

Proposition 2.1 *Under Assumptions 2.1, 2.2 and 2.3* $\beta_1(x) = \beta_2(x) = \beta(x)$ *is just-identified by the ratio*

$$\beta(x) = \frac{\mathbb{E}[Y_2|X = x] - \mathbb{E}[Y_1|X = x]}{x_2 - x_1} \tag{5}$$

for all $x \in \{x : x \in \mathcal{X},\ x_1 \neq x_2\}$.

Proof. Under Assumption 2.1 we have

$$\begin{aligned}
\mathbb{E}[Y_1|X] &= \alpha_1(X) + \beta_1(X)X_1 \\
\mathbb{E}[Y_2|X] &= \alpha_2(X) + \beta_2(X)X_2,
\end{aligned}$$

for $\alpha_t(X) = \mathbb{E}[a(A, U_t)|X]$ and $\beta_t(X) = \mathbb{E}[b(A, U_t)|X]$. Iterated expectations (which is allowable by part (ii) of Assumption 2.2), marginal stationarity (part (i) of Assumption 2.2) and time-invariance of A give

$$\beta_t(X) = \mathbb{E}[b(A, U_t)|X] = \mathbb{E}[\mathbb{E}[b(A, U_t)|X, A]|X] = \mathbb{E}\left[\widetilde{b}(X, A)\middle| X\right] = \beta(X),$$

for $\widetilde{b}(X, A) = \mathbb{E}[b(A, U_t)|X, A]$. This gives $\beta_1(X) = \beta_2(X) = \beta(X)$; a similar calculation gives $\alpha_t(X) = \mathbb{E}[a(A, U_t)|X] = \alpha(X)$. Taking differences across time periods and solving for $\beta(X)$ then gives (5). That $\beta(x)$ is just-identified follows directly from its definition as a conditional expectation function, linearity of Y_t in $a(A, U_t)$ and $b(A, U_t)$, and just-identification of $\mathbb{E}[Y_1|X]$ and $\mathbb{E}[Y_2|X]$. ■

To recover the APE we average $\beta(X)$ over the marginal distribution of X:

$$\gamma = \mathbb{E}[\beta(X)].$$

Since $\beta(x)$ is only identified on those points of the support of X for which $X_1 \neq X_2$ (i.e., for 'changers' or units which alter their choice of X_t across periods) we cannot, in general, calculate $\mathbb{E}[\beta(X)]$ without further assumptions. Consequently, unless all units change their value of X_t across periods, the APE is not identified. When X_t is discrete it is natural to construct bounds for γ or to compute the average of $\beta(X)$ among 'changers'. The latter approach is particularly simple and foreshadows our approach to estimation in the continuous case. When X_t is continuous we impose smoothness restrictions on $\beta(x)$ which are sufficient to point identify γ. We consider each case in turn.

Discrete regressor If $X_t \in \{0, \ldots, M\}$, then $\beta(x)$ is only identified for the $M(M+1)$ possible sequences of $x = (x_1, x_2)$ where $x_1 \neq x_2$. Although the APE is not identified, we can compute the average partial effect in the subpopulation of units who *change* their values of X_t across the two periods. Define this 'changers' average partial effect (CAPE) as

$$\gamma_t^C(x) = \mathbb{E}\left[b(A, U_t) \mid \Delta X \neq 0\right].$$

Let π_{ij} denote the probability of the event $X_1 = i$ and $X_2 = j$ with $i, j = 0, 1, \ldots, M$. The CAPE is defined as (invoking marginal stationarity):

$$\gamma^C = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbf{1}(i \neq j)\cdot\beta(i,j)\cdot\pi_{ij}}{\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbf{1}(i \neq j)\cdot\pi_{ij}}.$$

Under our maintained assumptions γ^C can be used to construct sharp bounds on γ using the general approach of Manski (2003).[5] Unfortunately, in many microeconometric applications neither bounds nor the CAPE will be particularly informative for the APE. In Card's (1996) analysis of the union wage premium, less than 10 percent of workers switch between collective bargaining coverage and non-coverage across periods (Table V, p. 971). In such cases γ^C is an average over a very particular population, while bounds on γ will be quite wide. When X_t is discrete, however, this is the very best we can do.

Before turning to the continuous case we briefly discuss identification of the LAR when X_t is discrete. As with the APE, the LAR is generally not identified. Instead we can identify the 'changers' local average response (CLAR). For $t = 1$ this is given by

$$\delta_1^C(m) = \frac{\sum_{i=1}^{M}\mathbf{1}(i \neq m)\cdot\beta(m,i)\cdot\pi_{mi}}{\sum_{i=1}^{M}\mathbf{1}(i \neq m)\cdot\pi_{mi}}, \qquad m = 0, \ldots, M-1.$$

Continuous regressor When X is continuous the set $\{x : x \in \mathcal{X}, \quad x_1 = x_2\}$ will generally be of measure zero. This suggests that, under mild smoothness conditions, $\beta(x)$ should be identifiable for all $x \in \mathcal{X}$. In particular, at those points where $x_1 = x_2$ We can then identify $\beta(x)$ by the limit

$$\beta(x_1, x_1) = \lim_{h \to 0}\frac{\mathbb{E}\left[Y_2 \mid X = (x_1, x_1 + h)\right] - \mathbb{E}\left[Y_1 \mid X = (x_1, x_1)\right]}{h}. \tag{6}$$

A sufficient condition for the above limit to exist is:

Assumption 2.4 (Smoothness) *$\beta(x)$ is continuous and differentiable in $\mathcal{X}$.*

[5] Chernozhukov, Fernández-Val, Hahn and Newey (2007) develop bounds for Y_t binary and $\mathbb{E}[Y_t \mid X, A] = \Phi(X_t'\beta + A)$. Their bounds exploit the parametric structure of the probit model.

Under this smoothness restriction we have the following Theorem.

Theorem 2.1 (IDENTIFICATION) *If X_t is continuously-valued and Assumptions 2.1, 2.2, 2.3 and 2.4 hold, then $\gamma_t(x_t) = \gamma$ and $\delta_t(x_t)$ are identified by*

$$\gamma = \mathbb{E}[\beta(X)], \qquad \delta_t(x_t) = \mathbb{E}[\beta(X)|\, X_t = x_t]$$

with $\beta(x)$ given by (5) or (6) as appropriate.

Proof. Straightforward and therefore omitted. ■

Observe that $\beta(x)$ is an average over the conditional distribution of (A, U_t) given X. Thus smoothness of $\beta(x)$ suggests that the distribution function of A given $X = x$ is smooth in x. Such smoothness conditions are often implied by correlated random effect specifications for A. A fixed effects purist could thus call our model (when X_t is continuous) a correlated random effects one. We maintain the fixed effects characterization because we view Assumption 2.4 as rather weak. In anycase estimation would be impossible without it.

2.1 Relationship to linear FE estimator

Our model can be used to provide a representation of the probability limit of the textbook FE estimator under misspecification.

$$\gamma^{FE} = \frac{\mathbb{E}[\Delta Y \Delta X]}{\mathbb{E}[\Delta X \Delta X]}.$$

Iterated expectations implies a representation of the numerator of γ^{FE} equal to

$$\begin{aligned}
\mathbb{E}[\Delta Y \Delta X] &= \sum_{k=-M}^{M} k \cdot \mathbb{E}[\Delta Y|\, \Delta X = k] \cdot \Pr(\Delta X = k) \\
&= \sum_{k=-M}^{M} k^2 \cdot \left\{ \frac{\sum_{i=0}^{M} \sum_{j=0}^{M} \mathbf{1}(j - i = k)\, \beta(i,j)\, \pi_{ij}}{\sum_{i=0}^{M} \sum_{j=0}^{M} \mathbf{1}(j - i = k) \cdot \pi_{ij}} \right\} \\
&\quad \times \left\{ \sum_{i=0}^{M} \sum_{j=0}^{M} \mathbf{1}(j - i = k) \cdot \pi_{ij} \right\} \\
&= \sum_{i=0}^{M} \sum_{j=0}^{M} \beta(i,j)\, \pi_{ij}\, (j - i)^2,
\end{aligned}$$

and a denominator representation of

$$\mathbb{E}[\Delta X \Delta X] = \sum_{k=-M}^{M} k^2 \cdot \left\{ \sum_{i=0}^{M} \sum_{j=0}^{M} \mathbf{1}(j - i = k) \cdot \pi_{ij} \right\} = \sum_{i=0}^{M} \sum_{j=0}^{M} \pi_{ij}\, (j - i)^2.$$

Combining these terms shows that the linear FE estimator is consistent for a weighted average partial effect

$$\gamma^{FE} = \sum_{i=0}^{M} \sum_{j=0}^{M} \beta(i,j)\,\omega_{ij}, \qquad \omega_{ij} = \frac{\pi_{ij}(j-i)^2}{\sum_{i=0}^{M}\sum_{j=0}^{M}\pi_{ij}(j-i)^2}. \tag{7}$$

This average gives zero weight to those units who do not change their values of X_t across periods and more weight (relative to their population frequency) to those units which make very large changes. When X_t is binary it is straightforward to show that $\gamma^{FE} = \gamma^{C}$, however, in general, the two estimands differ. Representation (7) is similar to the local average treatment effect (LATE) representation of the Wald-IV estimator's probability limit (Angrist, Imbens and Rubin 1996, Imbens 2007). An important difference is that 'changers', unlike 'compliers' in the LATE-context, can be direcly identified from the data. Consequently the weights in (7) are identified.

In independent work, Chernozhukov, Fernández-Val, Hahn and Newey (2007) obtain a related result in the context of a fixed effects probit model with a binary regressor (cf., Hahn 2001, LaPorte and Windmeijer 2005). Wooldridge (2005b), who maintains the CRC structure as we do, imposes the additional restriction (in our notation) that $\beta(i,j) = \mathbb{E}[b(A,U_t)]$ for $i \neq j$ (cf., Equation (14) on p. 387).[6] In that case equivalency of the FE probability limit and the APE follows directly.

2.2 Aggregate time effects

Marginal stationarity is a strong, albeit powerful, assumption. We can relax it slightly by allowing for aggregate time effects. Specifically we consider the model

$$Y_t = a_t(A,U_t) + b(A,U_t)\,X_t, \tag{8}$$

where $a_t(A,U_t)$, the mapping from A and U_t is time-specific.

Assumption 2.5 (Common Average Trends) $\mathbb{E}[\,a_2(A,U_t) - a_1(A,U_t)|\,X] = \delta$

Assumption 2.5 is an obvious generalization of the deterministic 'common trends' assumption routinely made in program evaluation studies (Heckman and Robb 1985, Meyer 1995, Angrist and Krueger 1999). In this model $\beta(x)$ is identified as long as some units do not change their values of X_t across periods. Such 'control units' identify the mean time effect, δ. Let x' denote a realization of X such that $x'_1 = x'_2$. Under (8) and our previous assumptions $\beta(x)$ is identified by the ratio

$$\beta(x) = \frac{\mathbb{E}[Y_2|\,X=x] - \mathbb{E}[Y_1|\,X=x] - \mathbb{E}[Y_2|\,X=x'] - \mathbb{E}[Y_1|\,X=x']}{x_2 - x_1} \tag{9}$$

for all $x \in \{x : x \in \mathcal{X},\ x_1 \neq x_2\}$. For $x_1 = x_2$ we can adapt expression (6) above. With $\beta(x)$ identified, identification of the APE and LAR follows directly.

Equation (9) can be used to provide a simple interpretation of the probability limit of the 'difference-in-differences' estimator under CRC misspecification (Card 1990, Meyer, Viscusi and

[6] Wooldridge (2005a) also assumes that the correlated random coefficients are time invariant.

Durbin 1995, Angrist and Krueger 1999). Let X_t denote the availability of some binary policy or treatment. In period one no units are exposed to the policy, while in period two some units are exposed, with the balance serving as a comparison group. Using (9) we can show that the probability limit of the difference-in-differences estimator is given by

$$\beta(0,1) = \mathbb{E}[Y_2 | X = (0,1)] - \mathbb{E}[Y_1 | X = (0,1)] - \mathbb{E}[Y_2 | X = (0,0)] - \mathbb{E}[Y_1 | X = (0,0)].$$

When the true model is of the form (8) and $\pi_{1j} = 0$ for $j = 0, 1$, then $\beta(0,1)$ equals the average treatment effect on the treated (ATT). The average treatment effect (ATE) is not identified under our assumptions. Athey and Imbens (2006) also generalize the textbook differences-in-differences model to allow for nonseparable heteroegeneity. Their assumptions are sufficient to identify the average treatment effect (ATE).

2.3 Connections to other research

Altonji and Matzkin (2005) also study semiparametric panel data models. They work with the general model given by (1) and the following exchangeability assumption:

Assumption 2.6 (Exchangeability) *(i)*

$$A, U_t | X_1, \ldots X_T \overset{D}{=} A, U_t | X_{p(1)}, \ldots X_{p(T)},$$

for $p(t) \in \{1, \ldots, T\}$, $p(t) \neq p(t')$, *(ii) the distribution of* (A, U_t) *given* X *is non-degenerate for all* $X \in \mathcal{X}$.

Observe that Assumption 2.6, unlike Assumption 2.2 above, *does* restrict the conditional distribution of A given X. Under Assumption 2.6 Altonji and Matzkin (2005, pp. 1062 - 3) show that the Fundamental Theorem of Symmetric Functions and the Weierstrass Approximation Theorem imply the distributional equality

$$A | X_1, \ldots X_T \overset{D}{=} A | \zeta_1(X), \ldots, \zeta_T(X),$$

where $\zeta_t(X)$ is the t^{th} elementary symmetric polynomial on X.[7] Because Assumption 2.6 is not sufficient to identify $\beta_t(x)$ Altonji and Matzkin (2005, pp. 1063 - 4) suggest either further restricting the conditional distribution of (A, U_t) given X or the form of the structural outcome equation.[8]

If we impose the CRC structure on (1), then Assumption 2.6 implies that

$$\begin{aligned} \mathbb{E}[Y_t | X] &= \alpha_t(X) + \beta_t(X) X_t \\ &= \alpha_t(\zeta_1(X), \zeta_2(X)) + \beta_t(\zeta_1(X), \zeta_2(X)) X_t, \end{aligned}$$

[7] These polynomials take the form $\zeta_1(X) = \sum_{1 \leq i \leq T} X_i$, $\zeta_2(X) = \sum_{1 \leq i < j \leq T} X_i X_j$, $\zeta_3(X) = \sum_{1 \leq i < j < k \leq T} X_i X_j X_k$, $\zeta_4(X) = \sum_{1 \leq i < j < k < l \leq T} X_i X_j X_k X_l$ to $\zeta_T(X) = \prod_{i=1}^{T} X_i$.

[8] One suggestion made by Altonji and Matzkin (2005) is to impose a correlated random coefficients structure on $m(X_t, A, U_t)$, as we do here (Equation immediately prior to Equation (2.6) on p. 1064).

for $t = 1, 2$.

Now consider x and x' such that $x_1 = x_2'$ and $x_2 = x_1'$ with $x_1 \neq x_2$ (i.e., x' is a permutation of x). It is easy to show that $\beta_t(x)$ is identified by

$$\beta_t(x) = \frac{\mathbb{E}[Y_t| X = x] - \mathbb{E}[Y_t| X' = x']}{x_t - x_t'}.$$

Exchangeability and the CRC structure are sufficient to identify $\beta_t(x)$ even if the outcome variable is only observed for a single period as along as X_t is observed in each period. Altonji and Matzkin (2005, p. 1065 - 66) argue that this feature of their approach is particularly attractive the context of sibling studies where the outcome (e.g., wages) may only be observed for a single older sibling, while the endogenous regressor (e.g., school quality) might be measured for younger as well as older siblings. In contrast, our approach requires that we observe Y_t in both periods.

Neither Assumption 2.2 or 2.6 nest the other. For example, while Assumption 2.2 does not restrict the conditional distribution of A given X it does exclude time-varying heteroscedasticty allowed by Assumption 2.6.

A natural combination of the two assumptions is:

Assumption 2.7 (Stationarity and Exchangeability) *(i)*

$$U_t| X, A \overset{D}{=} U_s| X, A, \qquad t \neq s,$$

(ii) the distribution of U_t *given* X *and* A *is non-degenerate for all* $(X, A) \in \mathcal{X} \times \mathcal{A}$, *(iii)*

$$A| X_1, \ldots X_T \overset{D}{=} A| X_{p(1)}, \ldots X_{p(T)},$$

for $p(t) \in \{1, \ldots, T\}$, $p(t) \neq p(t')$.

Under Assumption 2.7 $\beta(x)$ is overidentified since

$$\beta(x) = \frac{\mathbb{E}[Y_2| X = x] - \mathbb{E}[Y_1| X = x]}{x_2 - x_1} = \frac{\mathbb{E}[Y_2| X' = x'] - \mathbb{E}[Y_1| X' = x']}{x_2' - x_1'},$$

when x' is a permutation of x.

3 Estimation

In this section we discuss estimation of the "changers average partial effect" γ^C when the regressors are discrete, and the average partial effect γ when the regressors have a continuous distribution.

3.1 Discrete regressor, no aggregate time effect

When X_t is discrete a simple instrumental variables procedure can be used to estimate γ^C. Let $D_k = \mathbf{1}(\Delta X = k)$ and $\rho_k = \mathbb{E}[D_k]$. Let b_k be the solution to the population problem

$$\mathbb{E}\left[(\Delta Y - b_k \Delta X) D_k\right] = 0,$$

or

$$b_k = \frac{\mathbb{E}[\Delta X D_k'] \mathbb{E}[D_k D_k']^{-1} \mathbb{E}[D_k \Delta Y]}{\mathbb{E}\left[\Delta X D_k'\right] \mathbb{E}\left[D_k D_k'\right]^{-1} \mathbb{E}[D_k \Delta X']}.$$

Note that by iterated expectations and marginal stationarity,

$$\begin{aligned} \mathbb{E}[\Delta Y | \Delta X = k] &= \mathbb{E}[\mathbb{E}[\Delta Y | X, \Delta X = k] | \Delta X = k] \\ &= k\mathbb{E}[\beta(X) | \Delta X = k] \end{aligned}$$

when $k \neq 0$; this implies that

$$\begin{aligned} \mathbb{E}\left[D_k \Delta X'\right] &= \rho_k \cdot k \\ \mathbb{E}\left[D_k D_k'\right] &= \rho_k \\ \mathbb{E}[D_k \Delta Y] &= \rho_k \cdot k \cdot \mathbb{E}[\beta(X) | \Delta X = k], \end{aligned}$$

and hence that $b_k = \mathbb{E}[\beta(X) | \Delta X = k]$. Observe that for b_k to be well-defined we require that $k \neq 0$. Let $b = (b_{-M}, \ldots, b_{-1}, b_{+1}, \ldots, b_{+M})'$ and $\rho = \left(\rho_{-M}, \ldots, \rho_{-1}, \rho_{+1}, \ldots, \rho_{+M}\right)'$, we have

$$\gamma^C = \frac{b'\rho}{\iota'\rho},$$

where ι is a $2M$ column vector of ones. Note that $\widehat{\gamma}^C$ can be calculated by the 2SLS fit of $\widehat{b}$ on ι using $\widehat{\rho}$ as an instrument.

3.2 Continuous regressor, no aggregate time effects

When X_t is continuously distributed – or, more precisely, when ΔX is continuously distributed in a neighborhood of zero – and no aggregate time effects are present, then Theorem 2.1 implies that the average partial effect γ is identified as

$$\begin{aligned} \gamma &= E\left[\frac{\Delta\mu(\mathbf{X})}{\Delta X}\right] \\ &= E\left[\frac{\Delta\mu(\mathbf{X})}{\Delta X} | \Delta X \neq 0\right], \end{aligned}$$

where

$$\begin{aligned}
\mu_t(\mathbf{x}) &= E[Y_t|\mathbf{X}=\mathbf{x}], \\
\Delta\mu(\mathbf{x}) &\equiv \mu_2(\mathbf{x})-\mu_1(\mathbf{x}) \\
&= E[\Delta Y|\mathbf{X}=\mathbf{x}].
\end{aligned}$$

Given this latter expression and a random sample of size N of observations on (Y_1, Y_2, X_1, X_2), a naive estimator of the APE γ would be

$$\begin{aligned}
\tilde{\gamma} &\equiv \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\Delta Y_i}{\Delta X_i}\right) \\
&= \frac{\sum_{i=1}^{N} 1(\Delta X_i \neq 0)\left(\frac{\Delta Y_i}{\Delta X_i}\right)}{\sum_{i=1}^{N} 1(\Delta X_i \neq 0)}.
\end{aligned}$$

However, if ΔX has a positive density in a neighborhood of zero, this estimator will be inconsistent in general, since $\Delta Y/\Delta X$ will not have finite expectation (unlike $\beta(\mathbf{X}) = \Delta\mu(\mathbf{X})/\Delta X$, whose expectation exists by assumption). To ensure quadratic-mean convergence, we consider instead a "trimmed" estimator of the form

$$\begin{aligned}
\hat{\gamma} &= \frac{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N)\left(\frac{\Delta Y_i}{\Delta X_i}\right)}{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N)} \\
&\equiv \hat{\gamma}(h_N),
\end{aligned}$$

where h_N is a deterministic bandwidth sequence tending to zero as N tends to infinity. (An alternative consistent estimator would replace the denominator by the sample size N.)

The estimator $\hat{\gamma}$ – which would be consistent for γ^C when $\mathbf{X}$ has finite support – has asymptotic properties similar to a standard (uniform) kernel regression estimator for a one-dimensional problem. In particular, it is straightforward to verify that

$$Var(\hat{\gamma}) = O\left(\frac{1}{Nh_N}\right) \gg O\left(\frac{1}{N}\right),$$

so the rate of convergence is necessarily slower than $1/N$ when $h_N \to 0$. Assuming in addition that the bias of $\hat{\gamma}(h)$ is geometric in the bandwidth parameter h – that is

$$E\left[1(|\Delta X_i| > h_n)\left(\frac{\Delta Y_i}{\Delta X_i}\right) - \beta(\mathbf{X}_i)\right] = E\left[1(|\Delta X_i| \leq h_n)\beta(\mathbf{X}_i)\right] = O(h_n^p)$$

for some $p > 0$ (typically $p = 2$) – the fastest rate of convergence of $\hat{\gamma}$ to γ in quadratic mean will be achieved when the bandwidth sequence takes the form

$$h_N^* = h_0\ N^{-1/(2p+1)},$$

which yields

$$\begin{aligned}\hat{\gamma}(h_N^*) - \gamma &= O_p(N^{-p/(2p+1)}) \\ &\gg O_p(N^{-1/2}).\end{aligned}$$

While the bandwidth sequence h_N^* achieves the fastest rate of convergence for this estimator, the corresponding asymptotic normal distribution for $\hat{\gamma}(h_N^*)$ will be centered at a bias term involving the derivative of $E[\beta(\mathbf{X})|\Delta X = d]$ at $d = 0$. The estimator $\hat{\gamma}$ will have an asymptotic (normal) distribution centered at zero if the bandwidth h_N converges to zero faster than h_N^*; assuming

$$h_N = o(N^{-1/(2p+1)}),$$

routine application of Liapunov's CLT for triangular arrays yields the asymptotic distribution for $\hat{\gamma}$,

$$\sqrt{Nh_N}\,(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, \phi_0 \sigma_0^2),$$

where

$$\phi_0 \equiv \lim_{h \downarrow 0} \frac{\Pr\{|\Delta X| \leq h\}}{h}$$

is the density of $|\Delta X|$ at zero and

$$\begin{aligned}\sigma_0^2 &\equiv Var[\Delta Y|\Delta X = 0] \\ &= \lim_{h \downarrow 0} Var[\Delta Y \mid -h < \Delta X < h].\end{aligned}$$

Assuming $p = 2$, the asymptotic distribution of $\hat{\gamma}$ is thus the same as the asymptotic distribution of a (uniform) kernel regression estimator of $E[\Delta Y|\Delta X = 0]$.

Heuristically, since the sample average of $\Delta\mu(\mathbf{X})/\Delta X$ would converge to γ at a parametric ($\sqrt{N}$) rate when ΔX is continuously distributed at zero, the asymptotic precision of $\hat{\gamma}$ is dominated by the precision to which $E[\Delta Y|\Delta X = d]$ can be estimated in a neighborhood of zero. (Hmmm....)

3.3 Continuous regressor, aggregate time effect

When aggregate time effects are present, and the "common trends" condition (Assumption 2.5) holds, then equation (9) implies that the average partial effect γ is identified as

$$\begin{aligned}\gamma &= E\left[\frac{\Delta\mu(\mathbf{X}) - \delta}{\Delta X}\right] \\ &= E\left[\frac{\Delta\mu(\mathbf{X}) - \delta}{\Delta X}\,|\Delta X \neq 0\right],\end{aligned}$$

where now

$$\delta \equiv E[\Delta Y|\Delta X = 0].$$

If δ were known, a straightforward modification of the estimator proposed in the preceding section would be

$$\bar{\gamma} = \frac{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N) \left(\frac{\Delta Y_i - \delta}{\Delta X_i}\right)}{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N)},$$

which would inherit the large sample properties of $\hat{\gamma}$ above.

When δ is unknown, a natural counterpart to this infeasible estimator $\bar{\gamma}$ would be a uniform kernel estimator of δ,

$$\hat{\delta} \equiv \frac{\sum_{i=1}^{N} 1(|\Delta X_i| \leq h_N)\Delta Y_i}{\sum_{i=1}^{N} 1(|\Delta X_i| \leq h_N)},$$

whose asymptotic properties are well-known when ΔX is continuously distributed. Indeed, under standard regularity conditions a normalized version of $\hat{\delta}$ has the same asymptotic distribution as $\bar{\gamma}$,

$$\sqrt{Nh_N} \left(\hat{\delta} - \delta\right) \xrightarrow{d} N(0, \phi_0 \sigma_0^2),$$

where ϕ_0 and σ_0^2 are defined above. Furthermore, $\bar{\gamma}$ and $\hat{\delta}$ will be asymptotically independent, as the product of their influence functions will be zero by construction.

Given this estimator of the common trend δ, a feasible estimator of the APE γ would be

$$\hat{\gamma}_F = \frac{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N) \left(\frac{\Delta Y_i - \hat{\delta}}{\Delta X_i}\right)}{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N)}.$$

Though simple in appearance, derivation of the large-sample properties of $\hat{\gamma}_F$ is difficult, as its rate of convergence depends in a delicate way on the distribution of the regressors X. Writing the normalized version of $\hat{\gamma}_F$ in terms of its infeasible counterpart $\bar{\gamma}$ yields

$$\sqrt{Nh_N} \left(\hat{\gamma}_F - \gamma\right) = \sqrt{Nh_N} \left(\bar{\gamma} - \gamma\right) - \sqrt{Nh_N} \left(\hat{\delta} - \delta\right) \cdot \left[\frac{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N) \left(\frac{1}{\Delta X_i}\right)}{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N)}\right].$$

While the asymptotic behavior of the first two terms in this decomposition are straightforward, the rate of convergence of the third term,

$$\hat{\xi} \equiv \frac{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N) \left(\frac{1}{\Delta X_i}\right)}{\sum_{i=1}^{N} 1(|\Delta X_i| > h_N)},$$

will crucially depend upon the behavior of

$$\tau(d) \equiv E\left[sgn\{\Delta X\} \mid |\Delta X| \ = d\right]$$

for d in a neighborhood of zero.

If, for example, X_1 and X_2 are exchangeable, so that ΔX is symmetrically distributed about zero (at least for $|\Delta X|$ in a neighborhood of zero), then $\tau(d) \equiv 0$ and $\hat{\xi}$ will converge in probability to zero, ensuring the asymptotic equivalence of the feasible estimator $\hat{\gamma}_F$ and its infeasible counterpart $\bar{\gamma}$. Alternatively, if there is constant positive drift in the distribution of regressors, so that $\tau(d) = \tau(0) > 0$, then the third term $\hat{\xi}$ will diverge, with expectation of $O(\log(h_N^{-1}))$, which is $O(\log(N))$ if $h_N = O(N^{-r})$ for some $r > 0$. In the latter case, the asymptotic distribution of the feasible estimator $\hat{\gamma}_F$ will be dominated by the asymptotic distribution of the estimator $\hat{\delta}$ of the common trend. An intermediate case could have $\tau(d) = O(d)$ in a neighborhood of zero, with the third term converging in probability to some nonzero limit.

In any event, an asymptotic variance estimator for $\hat{\gamma}_F$ can be constructed if consistent estimators of the density ϕ_0 and conditional variance σ_0^2 terms appearing in the asymptotic variances of $\bar{\gamma}$ and $\hat{\delta}$ can be constructed. Under standard regularity conditions, the kernel estimators

$$\hat{\phi} \equiv \frac{1}{2Nh_N} \sum_{i=1}^{N} 1(|\Delta X_i| \leq h_N)$$

and

$$\hat{\sigma}^2 \equiv \frac{\sum_{i=1}^{N} 1(|\Delta X_i| \leq h_N) \left(\Delta Y_i - \hat{\delta}\right)^2}{\sum_{i=1}^{N} 1(|\Delta X_i| \leq h_N)}$$

should converge in probability to ϕ_0 and σ_0^2; given these estimators, an estimator of the asymptotic variance of the feasible estimator $\hat{\gamma}_F$ can be constructed as

$$\widehat{AV}(\hat{\gamma}_F) = \frac{\hat{\sigma}^2 \hat{\phi}}{Nh_N} \left(1 + \hat{\xi}^2\right),$$

for $\hat{\xi}$ the "third term" defined above. This estimator will automatically adapt to divergence of $\hat{\xi}$ or its convergence to a (possibly nonzero) constant in probability.

3.4 Mixed discrete-continuous regressors

In some applications the distribution of the regressors $\mathbf{X} = (X_1, X_2)$ may have mass points at a finite set of values, and will be continuously distributed elsewhere. If there is overlap in the mass points of X_1 and X_2, then the distribution of first differences ΔX will generally have a mass point at zero, and will otherwise be continuously distributed in a neighborhood of zero. In this setting, the average partial effect γ will generally differ from the "changers" counterpart γ^C, due to the nonzero probability that $\Delta X = 0$; while this mass point simplifies estimation of a nonzero common trend component δ (and the conditional variance of ΔY given $\Delta X = 0$), it complicates estimation of the APE γ because of its distinction from γ^C, which is the implicit estimand of $\hat{\gamma}$ and $\hat{\gamma}_F$ above.

When $\pi_0 \equiv \Pr\{\Delta X = 0\} > 0$, the estimator

$$\tilde{\delta} \equiv \frac{\sum_{i=1}^{N} 1(\Delta X_i = 0) \cdot \Delta Y_i}{\sum_{i=1}^{N} 1(\Delta X_i = 0)}$$

would clearly be a $\sqrt{N}$-consistent and asymptotically normal estimator for δ, as would be the (asymptotically equivalent) estimator $\hat{\delta}$, defined in the previous section, under standard regularity conditions. Using the decomposition for the feasible estimator $\hat{\gamma}_F$ of $\gamma^C \equiv E[\beta(\mathbf{X})|\Delta X \neq 0]$ in the previous section, it follows that

$$\begin{aligned}\sqrt{Nh_N}\left(\hat{\gamma}_F - \gamma^C\right) &= \sqrt{Nh_N}\left(\bar{\gamma} - \gamma^C\right) + O_p(\sqrt{h_N}) \cdot O_p(\log(h_n^{-1})) \\ &= \sqrt{Nh_N}\left(\bar{\gamma} - \gamma^C\right) + o_p(1),\end{aligned}$$

so that preliminary estimation of the common trend component δ will not affect the asymptotic distribution of the feasible estimator $\hat{\gamma}_F$. If a consistent estimator $\hat{\gamma}(0)$ of the "non-changers" effect

$$\gamma(0) \equiv E[\beta(\mathbf{X}) \mid \Delta X \ = 0]$$

can be constructed, a corresponding consistent estimator of the APE $\gamma = \pi_0\gamma(0) + (1-\pi_0)\gamma^C$ would be

$$\hat{\gamma} \equiv \hat{\pi}\hat{\gamma}(0) + (1 - \hat{\pi})\hat{\gamma}_F,$$

where

$$\hat{\pi} \equiv \frac{1}{N}\sum_{\iota=1}^{N} 1(|\Delta X| \leq h_N)$$

is a $\sqrt{N}$-consistent estimator for π_0.

Defining

$$\nu(d) \equiv E\left[\Delta Y \mid |\Delta X| \ = d\right],$$

the results of section 2 above imply that

$$\gamma(0) = \lim_{h \downarrow 0} \frac{\nu(h) - \nu(0)}{h};$$

thus, estimation of $\gamma(0)$ amounts to estimation of a (left) derivative at zero of the conditional mean of ΔY given $\Delta X = d$. One such consistent estimator would be the slope coefficient of a local linear regression of ΔY on a constant term and ΔX, i.e.,

$$\begin{pmatrix} \hat{\delta}(0) \\ \hat{\gamma}(0) \end{pmatrix} = \arg\min_{d,c} \sum_{i=1}^{n} 1(|\Delta X_i| \leq h_N) \cdot (\Delta Y_i - d - c\Delta X_i)^2,$$

with the intercept $\hat{\delta}(0)$ being an alternative ($\sqrt{N}$-)consistent estimator of the common trend δ. Since the rate of convergence of a nonparametric estimator of the derivative of a regression function is lower than for its level, the rate of convergence the combined estimator $\hat{\gamma} \equiv \hat{\pi}\hat{\gamma}(0) + (1-\hat{\pi})\hat{\gamma}_F$ of the APE γ will be the same as for $\hat{\gamma}(0)$, and the asymptotic distribution of the latter would dominate the asymptotic distribution of $\hat{\gamma}$ in this setting.

4 Extensions

4.1 Quantile partial effects

Knowledge of each quantile of $b(A, U_t)$, or the *quantile partial effects* (QPEs), allows for a complete characterization of the heterogeneous effect of exogenous changes in X_t on Y_t (cf., Koenker 2005, Chernozhukov and Hansen 2005, 2006, Imbens and Newey 2006). Formally, the QPE is defined as

$$\eta_t(\tau) \text{ is the } \tau^{th} \text{ quantile of } b(A, U_t). \tag{10}$$

In this subsection we show that, under an additional conditional 'comonotonicity' assumption on the random coefficients $\eta_t(\tau)$ is point identified for X_t continuous.

Assumption 4.1 (Conditional Comonotonicity) *(i)* $a(A, U_t) = a^*(V_t, X)$ *and* $b(A, U_t) = b^*(V_t, X)$ *for some scalar random variable* V_t *(ii)* $a^*(v_t, x)$ *and* $a^*(v_t, x)$ *are strictly increasing in* v_t*; (iii)* $X_t \in \mathcal{X} \subset \mathbb{R}^+$.

Assumption 4.1 imposes a strong form of conditional dependence across the random coefficients; the nature of which may be unrealistic in certain economic settings (Heckman, Smith and Clements 1997). Consider two individuals with a common history $X = x$. Individual one has a fixed effect of $A = a$ and a time-varying disturbance of $U_t = u_t$. The corresponding values for the second individual are a' and u'_t. Assumption 4.1 implies that if $a(a, u_t) > a(a', u'_t)$, then $b(a, u_t) > b(a', u'_t)$. This implies, that within subpopulations of units with a common history, the period-specific rank-order of absolute and comparative advantage is the same (cf., Jouini and Napp 2004). Chernozhukov and Hansen (2005, 2006) require a similar rank invariance assumption for their instrumental variables quantile treatment effects model (cf. Doksum 1974).

Let $F_{Y_1}^{-1}(\tau|\, x) = \inf\{y_1 \in \mathbb{R} : F_{Y_1}(\tau|\, x) \geq \tau\}$ and $F_{Y_2}^{-1}(\tau|\, x) = \inf\{y_2 \in \mathbb{R} : F_{Y_2}(\tau|\, x) \geq \tau\}$ give the τ^{th} quantiles of Y_1 and Y_2 in the subpopulation with history $X = x$. Under Assumption 4.1 we have (cf., Koenker 2005, pp. 59 - 62).

$$\begin{aligned} F_{Y_1}^{-1}(\tau|\, x) &= F_{a,1}^{-1}(\tau|\, x) + F_{b,1}^{-1}(\tau|\, x)\, x_1 \\ F_{Y_2}^{-1}(\tau|\, x) &= F_{a,2}^{-1}(\tau|\, x) + F_{b,2}^{-1}(\tau|\, x)\, x_2, \end{aligned}$$

where $F_{a,t}(\cdot|\, x)$ and $F_{b,t}(\cdot|\, x)$ are the conditional distribution functions of $a(A, U_t)$ and $b(A, U_t)$. Marginal stationarity further implies that $F_{a,1}(\cdot|\, x) = F_{a,2}(\cdot|\, x)$ and $F_{b,1}(\cdot|\, x) = F_{b,2}(\cdot|\, x)$. Imposing these restrictions and solving for $F_b^{-1}(\tau|\, x)$ gives (assuming $x_2 \neq x_1$):

$$F_b^{-1}(\tau|\, x) = \frac{F_{Y_2}^{-1}(\tau|\, x) - F_{Y_1}^{-1}(\tau|\, x)}{x_2 - x_1}.$$

Since $F_b^{-1}(\tau|\, x)$ is identified for all τ and x (using limits for x such that $x_1 = x_2$), $F_b(b|\, x)$ is also identified. We can therefore identify $\eta_t(\tau) = \eta(\tau)$ by averaging $F_b(b|\, x)$ over the marginal distribution of X to get $F_b(b)$ and then inverting the resulting distribution function.

Note that integrating over τ gives

$$\begin{aligned}\int F_b^{-1}(\tau|\,x)\,\mathrm{d}\tau &= \int \frac{F_{Y_2}^{-1}(\tau|\,x) - F_{Y_1}^{-1}(\tau|\,x)}{x_2 - x_1}\mathrm{d}\tau \\ &= \frac{\mathbb{E}[Y_2|\,x] - \mathbb{E}[Y_1|\,x]}{x_2 - x_1} \\ &= \beta(x).\end{aligned}$$

To incorporate aggregate time trends whilst still identifying QPEs requires a modification of Assumption 2.5 above.

Assumption 4.2 (Common Quantile Trends) $F_{a_2,t}^{-1}(\tau|\,x) - F_{a_1,t}^{-1}(\tau|\,x) = \delta(\tau)$.

Under Assumption 4.2 a result analogous to the mean case above with $F_b^{-1}(\tau|\,x)$ identified by

$$F_b^{-1}(\tau|\,x) = \frac{F_{Y_2}^{-1}(\tau|\,x) - F_{Y_1}^{-1}(\tau|\,x) - F_{Y_2}^{-1}(\tau|\,x') - F_{Y_1}^{-1}(\tau|\,x')}{x_2 - x_1},$$

for x' such that $x'_1 = x'_2$.

4.2 Endogenous regressors

In empirical work applications of first differencing is often combined with the method of two stage least squares. Differencing eliminates time-invariant correlated heterogeneity while 'instrumenting' deals with any remaining contemporaneous endogeneity. Wooldridge (2002, pp. 307 - 324) provides a textbook introduction to this method as well as several empirical examples. Murtazashvili and Wooldridge (2008) provide conditions under which the fixed effects instrumental variables (FE-IV) is consistent for the APE when the true model takes a CRC form. Alonji and Matzkin (2001) study the identifying power of exchangeability a triangular panel data model.

As above, we work with a structural outcome equation of

$$Y_t = a(A, U_t) + b(A, U_t)\,X_t.$$

However now we allow for the possibility of additional, contemporaneous, endogeneity. The first stage or *selection equation* is given by

$$X_t = g_t(W, V_t), \tag{11}$$

where $g_t(\cdot)$ is strictly monotone (normalized to be increasing) in its third argument. In this model endogeneity of X_t manifests itself in two ways: correlation between X_t and A and V_t and U_t.

Our first assumption is independence of V_t and W :

Assumption 4.3 (Independence):

$$V_t \perp W, \quad t = 1, 2.$$

As is well known, the mean independence version of Assumption 4.3 rules out V_t affecting future realizations of W_t or vice versa (Chamberlain 1984). One way to motivate (4.3) is through a correlated random effects model. Assume that

$$X_t = \widetilde{g}\left(W_t, B + V_t^*\right),$$

where B is a time-invariant effect, $\widetilde{g}(\cdot)$ is increasing in its second argument and V_t^* is independent of W. If we assume that $B = \lambda(W) + B^*$ with B^* independent of W (i.e., a correlated random effect), then we get (4.3) with $V_t = B^* + V_t^*$ and $g_t(W, V_t) = \widetilde{g}(W_t, \lambda(W) + V_t)$. If $\widetilde{g}(W_t, B + V_t^*) = \pi(W) + B + V_t^*$ we have the nonparametric panel data model with additive heterogeneity analyzed by Porter (1996) and Das (2003). Under Assumption 4.3 V_t is identified by $X_t - \mathbb{E}[X_t | W]$.

Our next identifying assumption is marginal stationarity of (U_t, V_t):

Assumption 4.4 (Marginal Stationarity): *(i)*

$$\left(U_t, V_t\right) | W, A \overset{D}{=} \left(U_s, V_s\right) | W, A, \qquad t \neq s.$$

(ii) the distribution of (U_t, V_t) given W and A is non-degenerate for all $(W, A) \in \mathcal{W} \times \mathcal{A}$.

Let $\beta_t(w, v_t) = \mathbb{E}[b(A, U_t) | W = w, V_t = v_t]$ be the average partial effect in the subpopulation of individuals with the instrument history $W = w$ and first-stage cost shock $V_t = v_t$. Under Assumptions 4.3 and 4.4 we have the following Proposition.

Proposition 4.1 *Under Assumptions 2.1, 4.3 and 4.4 $\beta_1(w, v_1) = \beta_2(w, v_1) = \beta(w, v_1)$ is just identified by the ratio*

$$\beta(w, v_1) = \frac{\mathbb{E}[Y_2 | W = w, V_2 = v_1] - \mathbb{E}[Y_1 | W = w, V_1 = v_1]}{x_2 - x_1} \tag{12}$$

$x_t = g_t(w, v_1)$ for $t = 1, 2$ and all $v_1 \in \mathcal{V}_1$ and $w \in \{w : w \in \mathcal{W},\ w_1 \neq w_2\}$.

Proof. Under Assumption 2.1 we have and the one-to-one mapping from X_t to V_t conditional on W implied by (11) we have

$$\begin{aligned}
\mathbb{E}[Y_t | W, X_t] &= \mathbb{E}[a(A, U_t) | W, X_t] + \mathbb{E}[b(A, U_t) | W, X_t] X_t \\
&= \mathbb{E}[a(A, U_t) | W, V_t] + \mathbb{E}[b(A, U_t) | W, V_t] X_t \\
&= \alpha_t(W, V_t) + \beta_t(W, V_t) X_t
\end{aligned}$$

for $\alpha_t(W, V_t) = \mathbb{E}[a(A, U_t) | W, V_t]$ and $\beta_t(W, V_t) = \mathbb{E}[b(A, U_t) | W, V_t]$. Iterated expectations (which is allowable by part (ii) of Assumption 4.4), marginal stationarity and time-invariance of A give

$$\beta_t(W, V_t) = \mathbb{E}[\mathbb{E}[a(A, U_t) | W, A, V_t] | W, V_t] = \mathbb{E}\left[\widetilde{b}(W, A, V_t) \middle| W, V_t\right] = \beta(W, V_t),$$

for $\widetilde{b}(W, A, V_t) = \mathbb{E}[b(A, U_t)| W, A, V_t]$. This gives $\beta_1(w, v_1) = \beta_2(w, v_1) = \beta(w, v_1)$; a similar calculation gives $\alpha_t(A, V_t) = \mathbb{E}[a(A, U_t)| A, V_t] = \alpha(A, V_t)$. Consider two subpopulations defined by their values of $W = (W_1', W_2')'$ and $V = (V_1', V_2')'$. In the first sub-population $W = w, V_1 = v_1, V_2 = v_2$, while in the second $W = w', V_1 = v_1', V_2 = v_2'$. Assume that $w = w'$ with $w_1 \neq w_2$ and $v_1 = v_2'$. From the structural outcome equation we have

$$\mathbb{E}[Y_2| W = w, V_2 = v_1] - \mathbb{E}[Y_1| W = w, V_1 = v_1] = \beta(w, v_1)\{g_2(w, v_1) - g_1(w, v_1)\}.$$

And from the selection equation we have

$$x_1 - x_2 = g_2(w, v_1) - g_1(w, v_1).$$

The ratio of these two differences gives (12). That $\beta(w, v_1)$ is just-identified follows directly from its definition as a conditional expectation function, linearity of Y_t in $a(A, U_t)$ and $b(A, U_t)$, and just-identification of $\mathbb{E}[Y_1| W, V_t]$ and $\mathbb{E}[Y_2| W, V_t]$. ■

4.2.1 Estimation with endogenous regressors

We propose to estimate $\beta(w, v_1)$ by

$$\widehat{\beta}(w, v_1) = \arg\min_b \begin{pmatrix} N \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} K_h(W_i - w, W_j - w, \widehat{V}_{1i} - v_1, \widehat{V}_{2j} - v_1) \times [(Y_{1i} - Y_{2j}) - b(X_{1i} - X_{2j})]^2,$$

where $\widehat{V}_{1i}$ and $\widehat{V}_{2i}$ are the residuals associated with the nonparametric regression fits of X_1 and X_2 on W.

This estimator is related to the one first proposed by Powell (1987, 2001). In closed form we have

$$\widehat{\beta}(w, v_1) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} K_h(W_i - w, W_j - w, \widehat{V}_{1i} - v_1, \widehat{V}_{2j} - v_1)(X_{1i} - X_{2j})(Y_{1i} - Y_{2j})}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} K_h(W_i - w, W_j - w, \widehat{V}_{1i} - v_1, \widehat{V}_{2j} - v_1)(X_{1i} - X_{2j})^2}.$$

We can then estimate the APE by the sample average

$$\widehat{\gamma} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\beta}(W_i, \widehat{V}_{1i}).$$

To get the LAR we use the partial mean (Newey 1994b).

$$\widehat{\delta}(x_t) = \frac{\sum_{i=1}^{N} K_h(X_{it} - x_t)\beta(W_i, X_{it} - \widehat{\mathbb{E}}[X_t| W_i])}{\sum_{i=1}^{N} K_h(X_{it} - x_t)}.$$

4.3 Multiple regressors and time periods

[To be completed]

5 Conclusion

[To be completed]

References

Abrevaya, Jason. (2000). "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics* 95 (1): 1 - 23.

Ahn, Hyungtaik and James L. Powell. (1993). "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics* 58 (1-2): 3 - 29.

Altonji, Joseph G. and Rosa L. Matzkin (2001). "Panel data estimators for nonseparable models with endogenous regressors," *NBER Technical Working Paper T0267.*

Altonji, Joseph G. and Rosa L. Matzkin (2005). "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica* 73 (4): 1053 - 1102.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. (1996). "Identification of causal effects using instrumental variables," *Journal of the American Statistical Association* 91 (434): 444 - 455.

Angrist, Joshua D. and Alan B. Krueger. (1999). "Empirical strategies in labor economics," *Handbook of Labor Economics* 3 (1): 1277 - 1366 (O. C. Ashenfelter & D. Card, Eds.). Amsterdam: North-Holland.

Angrist, Joshua D. and Adriana Kugler. (2008). Rural windfall or a new resource curse? Coca, income and civil conflict in Colombia," *Review of Economics and Statistics*, forthcoming.

Aradillas-Lopez, Andres, Bo Honoré and James L. Powell. (2007). "Pairwise difference estimation with nonparametric control variables," *Mimeo.*

Arellano, Manuel and Raquel Carrasco. (2003). "Binary choice panel data models with predetermined variables," *Journal of Econometrics* 115 (1): 125 - 157.

Arellano, Manuel and Bo Honoré. (2001). "Panel data models: some recent developments," *Handbook of Econometrics* 5: 3229 - 3298 (J. Heckman & E. Leamer, Eds.). Amsterdam: North-Holland.

Arellano, Manuel. (2003). "Discrete choices with panel data," *Investigacions Económicas* 27 (3): 423 - 458.

Ashenfelter, Orley and Alan Krueger (1994). "Estimates of the economic return to schooling from a new sample of twins," American Economic Review 84 (5): 1157 - 1173.

Athey, Susan and Guido W. Imbens. (2006). "Identification and inference in non-linear difference-in-differences models," *Econometrica* 74 (2): 431 - 497.

Besley, Timothy and Anne Case. (2000). "Unnatural experiments? Estimating the incidence of endogenous policies," Economic Journal 110 (467): F672 - F694.

Bester,C. Alan and Christian Hansen. (2007). "Identification of marginal effects in a nonparametric correlated random effects model," *Mimeo.*

Blundell, Richard W. and James L. Powell. (2003). "Endogeneity in nonparametric and semiparametric regression models," *Advances in Economics and Econometrics: Theory and Applications II*: 312 - 357. (M. Dewatripont, L.P. Hansen, S. J. Turnovsky, Eds.). Cambridge: Cambridge University Press.

Blundell, Richard W. and James L. Powell. (2004). "Endogeneity in semiparametric binary response models," *Review of Economic Studies* 71 (3): 655 - 679.

Browning, Martin and Jesus Carro. (2007). "Heterogeneity and microeconometrics modelling," *Advances in Economics and Econometrics: Theory and Applications III*: 47 - 74. (R. Blundell, W. Newey & T. Persson, Eds.). Cambridge: Cambridge University Press.

Burgess, Robin and Rohini Pande. (2005). "Can rural banks reduce poverty? Evidence from the Indian social banking experiment," *American Economic Review* 95 (3): 780 - 795.

Card, David. (1990). "The impact of the Mariel Boatlift on the Miami labor market," Industrial and Labor Relations Review 43 (2): 245 - 257.

Card, David. (1996). "The effect of unions on the structure of wages: a longitudinal analysis," *Econometrica* 64 (4): 957 - 979.

Card, David. (2001). "Estimating the return to schooling: progress on some persistent econometric problems," *Econometrica* 69 (5): 1127 - 1160..

Card, David and Alan B. Krueger. (1994). "Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania," *American Economic Review* 84 (4): 772 - 793.

Chamberlain, Gary. (1980). "Analysis of covariance with qualitative data," *Review of Economic Studies* 47 (1): 225 - 238.

Chamberlain, Gary. (1982). "Multivariate regression models for panel data," *Journal of Econometrics* 18 (1): 5 - 46.

Chamberlain, Gary. (1984). "Panel data," *Handbook of Econometrics 2:* 1247 - 1318 (Z. Griliches & M.D. Intriligator, Eds.). Amsterdam: North-Holland.

Chamberlain, Gary. (1992). "Binary response models for panel data: identification and information," *Mimeo.*

Chay, Kenneth Y. and Michael Greenstone. (2005). "Does air quality matter? Evidence from the housing market," *Journal of Political Economy* 113 (2): 376 - 424.

Chernozhukov, Victor, Jinyong Hahn and Whitney Newey. (2005). "Bounds analysis in panel models with correlated random effects," *Mimeo.*

Chernozhukov, Victor, Christian Hansen. (2005). "An IV model of quantile treatment effects," *Econometrica* 73 (1): 245 - 261.

Chernozhukov, Victor, Christian Hansen. (2006). "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics* 132 (2): 491 - 525.

Chesher, Andrew. (2007). "Identification of Nonadditive Structural Functions," *Advances in Economics and Econometrics: Theory and Applications III*: 1 - 16. (R. Blundell, W. Newey & T. Persson, Eds.). Cambridge: Cambridge University Press.

Currie, Janet and Duncan Thomas. (1995). "Does Head Start make a difference?" *American Economic Review* 85 (3): 341 - 364.

Das, Mitali. (2003). "Identification and sequential estimation of panel data models with insufficient exclusion restrictions," *Journal of Econometrics* 114 (2): 297 - 328.

Doksum, Kjell. (1974). "Empirical probability plots and statistical inference for nonlinear models in the two-sample case," *Annals of Statistics* 2 (2): 267 - 277.

Dulfo, Esther. (2001). "Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment," *American Economic Review* 91 (4): 795 - 813.

Finkelstein, Amy. (2002). "The effect of tax subsidies to employer-provided supplementary health insurance: evidence from Canada," *Journal of Public Economics* 84 (3): 305 - 339.

Florens, J. P., J. J. Heckman, C. Meghir and E. Vytlacil. (2007). "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogenous effects," *Mimeo.*

Griliches, Zvi. (1979). "Sibling models and data in economics: beginnings of a survey," *Journal of Political Economy* 87 (5): S37 - S64.

Hahn, Jinyong. (2001). "Comment: binary regressors in nonlinear panel-data models with fixed effects," *Journal of Business and Economics Statistics* 19 (1): 16 - 17.

Heckman, James J. and Richard Robb, Jr. (1985). "Alternative methods for evaluating the impact of interventions," *Longitudinal Analysis of Labor Market Data*: 156 - 245 (J.J. Heckman & B. Singer, Eds.). Cambridge: Cambridge University Press.

Heckman, James J., Jeffrey Smith and Nancy Clements. (1997). "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts," *Review of Economic Studies* 64 (4): 487 - 535.

Heckman, James and Edward Vytlacil. (1998). "Instrumental variables methods for the correlated random coefficient model: estimating the average rate of return to schooling when the return is correlated with schooling," *Journal of Human Resources* 33 (4): 974 - 987.

Honoré, Bo E. (1992). "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects," *Econometrica* 60 (3): 533 - 565.

Honoré, Bo E. and James L. Powell. (2005). "Pairwise difference estimators for nonlinear models," *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*: 520 - 553 (D.W.K. Andrews & J.H. Stock, Eds.). Cambridge: Cambridge University Press.

Honoré, Bo E. and Elie Tamer. (2006). "Bounds on parameters in panel dynamic discrete choice models," *Econometrica* 74 (3): 611 - 629.

Imbens, Guido W. (2007). "Nonadditive models with endogenous regressors," *Advances in Economics and Econometrics: Theory and Applications III*: 17 - 46. (R. Blundell, W. Newey & T. Persson, Eds.). Cambridge: Cambridge University Press.

Imbens, Guido W. and Whitney K. Newey. (2002). "Identification and estimation of triangular simultaneous equations models without additivity," *NBER Technical Working Paper 285.*

Imbens, Guido W. and Whitney K. Newey. (2006). "Identification and estimation of triangular simultaneous equations models without additivity," *Mimeo.*

Jouini, Elyés and Clotilde Napp. (2004). "Conditional comonotonicity," *Decisions in Economics and Finance* 27 (2): 153 - 166.

Koenker, Roger. (2005). *Quantile Regression.* Cambridge: Cambridge University Press.

Kyriazidou, Ekaterini. (1997). "Estimation of a panel data sample selection model," Econometrica 65 (6): 1335 - 1364.

LaPorte, Audrey and Frank Windmeijer. (2005). "Estimation of panel data models with binary indicators when treatment effects are not constant over time," *Economics Letters* 88 (3): 389 - 396.

Lemieux, Thomas and David Card. (2001). "Education, earnings and the 'Canadian G.I. Bill'," *Canadian Joural of Economics* 34 (2): 313 - 344.

Manski, Charles F. (1987). "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica* 55 (2): 357 - 362.

Manski, Charles F. (1995). *Identification Problems in the Social Sciences.* Cambridge, MA: Harvard University Press.

Manski, Charles F. (1997). "Monotone treatment response," *Econometrica* 65 (6): 1311 - 1334.

Meyer, Bruce D. (1995). "Natural and quasi-experiments in economics," *Journal of Business and Economic Statistics* 13 (2): 151 - 161.

Meyer, Bruce D., W. Kip Viscusi and David L. Durbin. (1995). "Workers' compensation and injury duration: evidence from a natural experiment," American Economic Review 85 (3): 322 - 340.

Mundlak, Yair. (1961). "Empirical production function free of management bias," *Journal of Farm Economics* 43 (1): 44 - 56.

Mundlak, Yair. (1978). "On the pooling of time series and cross section data," *Econometrica* 46 (1): 69 - 85.

Murtazashvili, Irina and Jeffrey M. Wooldridge. (2008). "Fixed effects instrumental variables estimation in correlated random coefficient panel data models," *Journal of Econometrics*142 (1): 539 - 552.

Porter, Jack. R. (1996). *Essays in Econometrics.* PhD diss., Massachusetts Institute of Technology.

Powell, James L. (2001). "Semiparametric estimation of bivariate latent variables models," *SSRI Working Paper No. 8704.* Madision, WI: Department of Economics, University of Wisconsin.

Powell, James L. (2001). "Semiparametric estimation of censored selection models," *Nonlinear Statistical Modelling* (C. Hsiao et al., Eds.). Cambridge: Cambridge University Press.

Powell, James L, James H. Stock and Thomas M. Stoker. (1989). "Semiparametric estimation of index coefficients," *Econometrica* 57 (6): 1403 -1430.

Newey, Whitney K. (1994a). "The asymptotic variance of semiparametric estimators," *Econometrica* 62 (6): 1349 - 1382.

Newey, Whitney K. (1994b). "Kernel estimation of partial means and a general variance estimator," *Econometric Theory* 10 (2): 233 - 253.

Newey, Whitney K., James L. Powell and Francis Vella. (1999). "Nonparametric estimation of triangular simultaneous equation models," *Econometrica* 67 (3): 565 - 603.

Todd, Petra E. and Kenneth I. Wolpin. (2008). "The production of cognitive achievement in children: home, school and racial test-score gaps," *Journal of Human Capital* 1 (1): X-X.

Wooldridge, Jeffrey M. (2005a). “Unobserved heterogeneity and estimation of average partial effects,” *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*: 27 - 55 (D.W.K. Andrews & J.H. Stock, Eds.). Cambridge: Cambridge University Press.

Wooldridge, Jeffrey M. (2005b). “Fixed-effects and related estimators for correlated-random coefficient and treatment-effect panel data models,” *Review of Economics and Statistics* 87 (2): 385 - 390.