Identification and estimation of ‘irregular’ correlated random coefficient models[1]

Bryan S. Graham† and James L. Powell◇

Initial Draft: December 2007

This Draft: October 15, 2008

Abstract

In this paper we study identification and estimation of the causal effect of a small change in an endogenous regressor on a continuously-valued outcome of interest using panel data. We focus on the average partial effect (APE) over the full population distribution of unobserved heterogeneity (e.g., Chamberlain, 1984; Blundell and Powell, 2003; Wooldridge, 2005a). In our basic model the outcome of interest varies linearly with a (scalar) regressor, but with an intercept and slope coefficient that may vary across units and over time in a way which depends on the regressor. This model is a special case of Chamberlain's (1980b, 1982, 1992a) correlated random coefficients (CRC) model, but not does not satisfy the regularity conditions he imposes. Irregularity, while precluding estimation at parametric rates, does not result in a loss of identification under mild smoothness conditions. We show how two measures of the outcome and regressor for each unit are sufficient for identification of the APE as well as aggregate time trends. We identify aggregate trends using units with a zero first difference in the regressor or, in the language of Chamberlain (1980b, 1982), ‘stayers’ and the average partial effect using units with non-zero first differences or ‘movers’. We discuss extensions of our approach to models with multiple regressors and more than two time periods. We use our methods to estimate the average elasticity of calorie consumption with respect to total outlay for a sample of poor Nicaraguan households (cf., Strauss and Thomas, 1995; Subramanian and Deaton, 1996). Our CRC average elasticity estimate declines with total outlay more sharply than its parametric counterpart.

JEL Classification: C14, C23, C33

Key Words: Panel Data, Correlated Random Coefficients, Average Partial Effects, Kernel Regression, Calorie Demand

---

[1] We would like to thank seminar participants at UC - Berkeley, UCLA, USC and Harvard, members of the Berkeley Econometrics Reading Group and participants in the Conference in Economics and Statistics in honor of Theodore W. Anderson's 90th Birthday (Stanford University), the Copenhagen Microeconometrics Summer Workshop and the JAE Conference on Distributional Dynamics (CEMFI, Madrid) for comments and feedback. Discussions with Manuel Arellano, Stéphane Bonhomme, Gary Chamberlain, Bo Honoré, Michael Jansson and Edward Vytlacil were helpful in numerous ways. Max Kasy provided research assistance. 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 linear regression models is among the most widespread methods of 'controlling for' omitted variables in empirical work (e.g., Griliches, 1979; Currie and Thomas, 1995; Card, 1996; Altonji and Dunn, 1996). The appropriateness of this modelling strategy, however, hinges on any time-invariant correlated heterogeneity entering the outcome equation additively. Unfortunately, additivity, while statistically convenient, is difficult to motivate economically (cf., 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.

In this paper we study the use of panel data for identifying and estimating what is arguably the simplest statistical model admitting nonseparable heterogeneity: the *correlated random coefficients* (CRC) model. 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. In the most basic model we analyze the structural outcome equation is given by

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

where Y_t is a scalar continuously-valued outcome of interest, X_t a scalar choice variable, A time-invariant unobserved unit-level heterogeneity and U_t a time-varying disturbance. Both A and U_t may be vector-valued. The functions $a_t\left(A, U_t\right)$ and $b_t\left(A, U_t\right)$, which we allow to vary over time (albeit in a restricted way), map the the time-invariant and time-varying heterogeneity into unit-by-period-specific intercept and slope coefficients.

Equation (1) is structural in the sense that the *unit-specific* function

$$Y_t\left(x_t\right) = a_t\left(A, U_t\right) + b_t\left(A, U_t\right) x_t \tag{2}$$

traces out a unit's period t potential outcome under different hypothetical values of x_t.[3] Equation (2) differs from the the textbook linear panel data model (with unit-specific intercepts, but otherwise constant regressor coefficients) in that the effect of a small change in x_t generally varies across units and/or time.

We study estimands which characterize the effect on an exogenous change in X_t on the probability distribution of Y_t. For concreteness we focus on identification and estimation of the *average partial effect* (APE) (cf., Chamberlain, 1984; Blundell and Powell, 2003; Wooldridge, 2005a. Developing parallel results for the *local average response* (LAR), as in Altonji and Matzkin (2005) and Bester and Hansen (2007) appears to be straightforward.. In the binary regressor case these two objects correspond to the average treatment effect (ATE) and the average treatment effect on the

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

[3] Throughout we use capital letters to denote random variables and lower case letters specific realizations of them.

treated (ATT) (cf., Florens, Heckman, Meghir and Vytlacil, 2008).

The average partial effect is given by

$$\beta_t \equiv \mathbb{E}\left[\frac{\partial Y_t(x_t)}{\partial x_t}\right] = \mathbb{E}\left[b_t(A, U_t)\right]. \tag{3}$$

Because of linearity of (2), β_t does not depend on x_t.[4]

Identification and estimation of (3) is nontrivial because X_t may vary systematically with A and/or U_t. To see the consequences of such dependence observe that the derivative of the regression function of Y_t given $X = (X_1, \ldots, X_T)'$ does not identify a structural parameter. Differentiating through the integral we have

$$\frac{\partial \mathbb{E}\left[Y_t | X = x\right]}{\partial x_t} = \beta_t(x) + \mathbb{E}\left[Y_t(X_t)\mathbb{S}_X(A, U_t | X) | X = x\right]$$

with $\beta_t(x) = \mathbb{E}\left[b_t(A, U_t) | X = x\right]$ and $\mathbb{S}_X(A, U_t | X) = \nabla_X \log f(A, U_t | X)$. The second term is what Chamberlain (1982) calls heterogeneity bias. If the (log) density of the unobserved heterogeneity varies sharply with x_t – corresponding to 'selection bias' or 'endogeneity' in a unit's choice of x_t – then this type of bias can be quite large.

To contextualize our contributions within the wider panel data literature it is useful to consider the more general outcome response function:

$$Y_t(x_t) = m(x_t, A, U_t).$$

Identification of the APE in the above model may be achieved by one of two main classes of restrictions. The *correlated random effects* approach invokes smoothness priors on the joint distribution of $(U, A)| X$; with $U = (U_1, \ldots, U_T)'$ and $X = (X_1, \ldots, X_T)'$. Mundlak (1978a,b) and Chamberlain (1980a, 1984) develop this approach for the case where $m(X_t, A, U_t)$ and $F(U, A| X)$ are parametrically specified. Newey (1994a) considers a semiparametric specification for $F(U, A| X)$ (cf., Arellano and Carrasco 2003). Recently, Altonji and Matzkin (2005) have extended this idea to the case where $m(X_t, A, U_t)$ is either semi- or non-parametric along with $F(U, A| X)$ (cf., Bester and Hansen 2007).

The *fixed effects* approach imposes restrictions on $m(X_t, A, U_t)$ and $F(U| X, A)$, while leaving $F(A| X)$, the distribution of the time-invariant heterogeneity, the so-called 'fixed effects', unrestricted. Chamberlain (1980a, 1984, 1992), Manski (1987), Honoré (1992) and Abrevaya (2000) are examples of this approach. Depending on the form of $m(X_t, A, U_t)$, the fixed effect approach may not allow for a complete characterization of the effect of exogenous changes in X_t on the probability distribution of Y_t. Instead only certain features of this relationship may be identified (e.g., ratios of the average partial effect of two regressors) (cf., Chamberlain, 1984; 1992b; Arellano and Honoré, 2001; Arellano 2003).

[4] If X_t is itself a function of a lower-dimensional choice variable R_t, the APE, defined in terms of r_t, may vary with r_t. Extending our results to this case is straightforward and we use such a formulation in the empirical application.

Our methods are of the 'fixed effect' variety. In addition to assuming the CRC structure for $Y_t(x_t)$ we impose a marginal stationarity restriction on $F(U_t|X, A)$, a restriction also used by Manski (1987), Honoré (1992) and Abrevaya (2000), however, other than some weak smoothness conditions, we leave $F(A|X)$ nonparametric.

Motivated by heterogeneity in the labor market returns to schooling, Card (1995, 2001) and Heckman and Vytlacil (1998) have studied identification and estimation of the CRC model using cross section data and 'instrumental variables' (cf., Garen, 1984; Heckman and Robb, 1985; Wooldridge, 1997, 2001, 2005a). This work belongs to larger body of research on nonparametric triangular systems (e.g., Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996; Heckman and Vytlacil, 2001; Blundell and Powell, 2003; Imbens and Newey, 2007; Florens, Heckman, Meghir and Vytlacil, 2008).[5]

The value of panel data for identification and estimation in the CRC model is comparatively less well understood. Mundlak (1961), while primarily focusing on a constant coefficients linear panel data model with unit-specific intercepts, briefly, and verbally, refers to the CRC model (p. 45).[6] In later work he studies estimators based on parametric specifications of the mean and variance of the random coefficients given all leads and lags of the regressors (Mundlak, 1978b).

The first analysis of CRC model that rigorously addresses identification issues in the context of panel data appears in Chamberlain (1980b, 1982). In later work Chamberlain (1992a, pp. 579 - 585) proposed an efficient method-of-moments estimator for the APE (cf., Wooldridge, 1999). Despite its innovative nature, and contemporary relevance given the resurgence of interest in models with heterogenous marginal effects, Chamberlain's work on the CRC model appears to have been largely underappreciated. For example, the CRC specification is not discussed in Chamberlain's own *Handbook of Econometrics* chapter (Chamberlain, 1984), while the panel data portion of Chamberlain (1992a) is only briefly reviewed in the more recent survey by Arellano and Honoré (2001).

The estimator proposed by Chamberlain (1992a) requires strong regularity conditions which, as we discuss further below, rule out substantively important economic models. Our contribution is to provide identification results, a consistent estimator and distribution theory for the case where Chamberlain's (1992a) information bound for the APE is zero. Singularity of the relevant information bound rules out estimation at parametric rates, nevertheless we show that consistent estimation at one-dimensional non-parametric rates is possible and, via empirical application, feasible. Interestingly irregularity also creates new identification opportunities, allowing us, for example, to identify aggregate time effects.

Wooldridge (2005b) also analyzes a CRC panel data model. His focus is on providing conditions under which the usual linear fixed effects (FE) estimator is consistent despite the presence of correlated random coefficients (cf., Chamberlain, 1982, p. 11). Fernández-Val (2005) develops bias correction methods for the CRC model in a large-N, large-T setting. Altonji and Matzkin

[5] Much of this research is surveyed by Imbens (2007).

[6] The exact reference is "The key to the estimation of the [average] slope of the infrafirm function is have at least two points of data on each f_i. In this case it is possible to get the slope of each of the lines f_i, average them and get the final estimate. That requires a combination of time series and cross section data."

(2005) and Bester and Hansen (2007) have developed new methods for using panel data to control for nonseparable unobserved heterogeneity. As their approaches are of the random effects variety, while our's are of the fixed effect variety, we view our methods as complementary to theirs.

Chamberlain (1982) showed that when X_t is discretely valued the APE is generally not identified (p. 13). However, Chernozhukov, Fernández-Val, Hahn and Newey (2008), working with more general forms for $\mathbb{E}\left[Y_t | X, A\right]$, show that when Y_t has bounded support the APE is partially identified and propose a method of estimating the identified set.[7] In contrast, in our setup the APE is point identified when X_t is continuously-valued. In fact, we are able to provide a characterization of when this estimand is semiparametrically just-identified. In that sense, our maintained assumptions are minimally sufficient (although not necessary).[8]

Porter (1996) and Das (2003) study nonparametric estimation of panel data model with additive unobserved heterogeneity. Honoré (1992) and Abrevaya (2000) consider models with nonseparable heterogeneity but, like Manski (1987), only identify index coefficients, not the APE. Arellano and Bonhomme (2008) also study identification in Chamberlain's (1992a) CRC model. Their focus is on identification and estimation of higher-order moments of the distribution of the random coefficients. Unlike us, they maintain Chamberlain's (1992a) regularity conditions as well as impose additional assumptions.

The next section reports identification results for the APE in a two period version of our core model. When X_t is discretely-valued our assumptions generally only bound the APE (appropriately defined to account for the discreteness of X_t). Our analysis of the discrete regressor case suggests useful interpretations of the probability limits of the linear fixed effects (FE) estimator and the 'difference-in-differences' (DID) estimator of the program evaluation literature (Card, 1990; Meyer, 1995; Angrist and Krueger, 1999; Athey and Imbens, 2006). When X_t is continuously valued, the case we focus upon, the APE is point identified. We also contrast our 'fixed effects' approach to identification with the semiparametric random effects methods developed by Altonji and Matzkin (2005).

Section 3 details our estimation approach. We begin with a discussion of the two period case. We then introduce a general multiple period CRC model and discuss its estimation. In that section we also relate our results to those of Chamberlain (1992a). Under Chamberlain's (1992a) conditions, which are not satisfied in our leading example, the APE is estimable at parametric rates. In contrast, our estimator has asymptotic properties similar to a standard one-dimensional kernel regression problem. This is a manifestation of the 'irregularity' of our model. In Section 4 discuss extensions of our approach to models with multiple regressors and more than two time periods. In that section we also compare our estimator with than of Chamberlain (1992a).

In Section 5, we use our methods to estimate the average elasticity of calorie demand with respect to total household resources in a sample of poor rural communities in Nicaragua. Our

[7] They consider the probit and logit models with unit-specific intercepts (in the index) in detail. They show how to construct bounds on the APE despite the incidental parameters problem (cf., Hahn, 2001) and provide conditions on the distribution of X_t such that these bounds shrink as T grows.

[8] Chesher (2007) provides an extended discussion of the value of 'just identifying' semiparametric restrictions.

sample is drawn from a population that participated in a pilot of the conditional cash transfer program Red de Protección Social (RPS). Hunger, conventionally measured, is widespread in the communities from which our sample is drawn; we estimate that immediately prior to the start of the RPS program over half of households had less then the required number of calories needed for all their members to engage in 'light activity' on a daily basis.[9]

Worldwide, the Food and Agricultural Organization (FAO) estimates that 854 million people suffered from protein-energy malnutrition in 2001-03 (FAO, 2006). Halving this number by 2015, in proportion to the world's total population, is the first United Nations Millennium Development Goal. Chronic malnutrition, particularly in early childhood, may adversely affect cognitive ability and economic productivity in the long-run (e.g., Dasgupta, 1993; Grantham-McGregor and Baker-Henningham, 2005; Case and Paxson, 2006; Hoddinott et al., 2008). A stated goal of the RPS program is to reduce childhood malnutrition, and consequently increase human capital, by directly augmenting household income in exchange for regular school attendance and participation in preventive health care check-ups.

The efficacy of this approach to reducing childhood malnutrition largely depends on the size of the average elasticity of calories demanded with respect to income across poor households.[10] While most estimates of the elasticity of calorie demand are significantly positive, several recent estimates are small in value and/or imprecisely estimated, casting doubt on the value of income-oriented anti-hunger programs (Behrman and Deolalikar, 1987; Strauss and Thomas, 1995; Subramanian and Deaton, 1996; Hoddinott, Skoufias and Washburn, 2000). Wolfe and Behrman (1983), using data from pre-revolutionary Nicaragua, estimate a calorie elasticity of just 0.01. Their estimate, if accurate, suggests that the income supplements provided by the RPS program should have little effect on caloric intake.

Disagreement about the size of the elasticity of calorie demand has prompted a vigorous methodological debate in development economics. Much of this debate has centered, appropriately so, on issues of measurement and measurement error (e.g., Behrman and Deolalikar, 1987; Bouis and Haddad, 1992; Bouis, 1994; Subramanian and Deaton 1996). The implications of household-level correlated heterogeneity in the underlying elasticity for estimating its average, in contrast, have not been examined. If, for example, a households' food preferences, or preferences towards child welfare, co-vary with those governing labor supply, then its elasticity will be correlated with total household resources. An estimation approach which presumes the absence of such heterogeneity will generally be inconsistent for the parameter of interest. Our statistical model and corresponding estimator provides an opportunity, albeit in a specific setting, for assessing the relevance these types of heterogeneities.

[9] We use Food and Agricultural Organization (FAO, 2001) gender- and age-specific energy requirements for 'light activity', as reported in Appendix 8 of Smith and Subandoro (2007), and our estimates of total calories available at the household-level to calculate the fraction of households suffering from 'food insecurity'. This approach to measuring food insecurity is not without its critics (e.g., Edmundson and Sukhatme 1990). Ferro-Luzzi (2005) provides a historical and conceptual overview of FAO/WHO food energy recommendations.

[10] Another motivation for studying this elasticity has to do with its role in theoretical models of nutrition-based poverty traps (e.g., Mirlees, 1975; Stiglitz, 1976; Bliss and Stern, 1978; Dasgupta and Ray, 1986, 1987).

We compare our CRC estimates of the elasticity of calorie demand with those estimated using standard panel data estimators (e.g., Behrman and Deolalikar, 1987; Bouis and Haddad, 1992), as well as those derived from the cross-sectional nonparametric regression techniques as in Subramanian and Deaton (1996), Strauss and Thomas (1995) and others. While the evidence is far from conclusive, we find that our CRC estimates of the average elasticity are higher at low-incomes, and lower at high-incomes, than those estimated by both of these alternative methods.

Section 5.5 summarizes, discusses a few simple extensions of our results and suggests areas for further research.

2 Identification: the two period case with a scalar regressor

We illustrate each of our main identification results for the case where X_t is scalar and $T = 2$. We generalize to panels are arbitrary length and multiple regressors in Section 3 below. Our first assumption is that the data generating process takes a correlated random coefficients form.

Assumption 2.1 (Correlated Random Coefficients)

$$Y_t = a_t (A, U_t) + b_t (A, U_t) X_t.$$

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

Assumption 2.2 (Marginal Stationarity) *(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* $(X, A) \in \mathcal{X} \times \mathcal{A}$.

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).

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 $X = (X_1, X_2)'$ and

$$\beta_t (x) \equiv \mathbb{E} [b_t (A, U_t) | X = x]$$

denote the average period t marginal effect of a change in x_t within the subpopulation of units where $X = x = (x_1, x_2)'$. Observe that $\beta_t (x)$ gives the average effect within a subpopulation defined by a common complete *history* of choices for X_t.

Below we discuss how to incorporate aggregate time effects into our analysis. However, for clarity of exposition, we begin by also invoking the additional restriction (which we relax below).

Assumption 2.4 (No Time Effects) *$a_1(a,u_1) = a_2(a,u_1)$ and $b_1(a,u_1) = b_2(a,u_1)$ for all $a \times u_1 \in \mathcal{A} \times \mathcal{U}$.*

Our first result shows that $\beta_t(x)$ is just-identified when $x_1 \neq x_2$.

Proposition 2.1 *Under Assumptions 2.1 to 2.4 $\beta_1(x) = \beta_2(x) = \beta(x)$ is just-identified by the ratio*

$$\beta(x) = \frac{\mathbb{E}[Y_2|X=x] - \mathbb{E}[Y_1|X=x]}{x_2 - x_1} \tag{4}$$

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

Proof. Under Assumptions 2.1 and 2.3 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 $\alpha_t(X) = \mathbb{E}[a_t(A,U_t)|X]$ and $\beta_t(X) = \mathbb{E}[b_t(A,U_t)|X]$. Iterated expectations (which is allowable by part (ii) of Assumption 2.2), marginal stationarity (part (i) of Assumption 2.2), time-invariance of A and Assumption 2.4 give

$$\beta_t(X) = \mathbb{E}[b_t(A,U_t)|X] = \mathbb{E}[\mathbb{E}[b_t(A,U_t)|X,A]|X] = \mathbb{E}[\widetilde{b}_t(X,A)\big|X] = \beta(X),$$

for $\widetilde{b}_t(X,A) = \mathbb{E}[b_t(A,U_t)|X,A]$. This gives $\beta_1(X) = \beta_2(X) = \beta(X)$; a similar calculation gives $\alpha_t(X) = \mathbb{E}[a_t(A,U_t)|X] = \alpha(X)$. Taking differences across time periods and solving for $\beta(X)$ then gives (4). That $\beta(x)$ is just-identified follows directly from its definition as a conditional expectation function, linearity of Y_t in $a_t(A,U_t)$ and $b_t(A,U_t)$, and just-identification of $\mathbb{E}[Y_1|X]$ and $\mathbb{E}[Y_2|X]$. ■

To recover the APE, which under Assumption 2.4 is constant over time, we average $\beta(X)$ over the marginal distribution of X:

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

Since $\beta(x)$ is only identified on those points of the support of X for which $X_1 \neq X_2$ (i.e., for 'movers' or units which alter their choice of X_t across periods) we cannot, in general, calculate $\mathbb{E}[\beta(X)]$ without further assumptions (Chamberlain 1982, p. 13). 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 β or to compute the average of $\beta(X)$ among 'movers'. The former idea is developed by Chernozhukov, Fernández-Val, Hahn and Newey (2008) is some generality. The latter approach, originally suggested by Chamberlain (1980b, 1982), is particularly simple and we review it since it 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 β. We consider each case in turn.

Discrete regressor If $X_t \in \{0, \dots, M\}$, 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 (Chamberlain 1980b, 1982). Define, invoking marginal stationarity and the absence of time effects, the 'movers' average partial effect (MAPE) as

$$\beta^M \equiv \mathbb{E}\left[b_t(A, U_t) \middle| \Delta X \neq 0 \right] = \frac{\mathbb{E}\left[\mathbf{1}(\Delta X \neq 0)\beta(X)\right]}{\mathbb{E}\left[\mathbf{1}(\Delta X \neq 0)\right]}. \tag{5}$$

Expression (5) is implicit in Chamberlain (1982, p. 13) who also noted that we have no information on $\beta^S = \mathbb{E}\left[b_t(A, U_t) \middle| \Delta X = 0 \right]$, or the 'stayers' average partial effect (SAPE). The data are consistent with β^S taking on any feasible value. When Y is continuously-valued along the real line, then any value for $\beta = \mathbb{E}[\beta(X)]$ is consistent with any given value for β^M. However, if Y has bounded support then β^M can be used to construct sharp bounds on β using the general approach of Manski (2003) as shown by Chernozhukov, Fernández-Val, Hahn and Newey (2008).

Many microeconometric applications are characterized by a preponderance of stayers. In Card's (1996) analysis of the union wage premium, for example, less than 10 percent of workers switch between collective bargaining coverage and non-coverage across periods (Table V, p. 971). In such cases β^M is an average over a very particular population, while bounds on β will be quite wide. When X_t is discrete, however, this is the very best we can do without invoking additional assumptions.

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

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

A sufficient condition for the above limit to exist is:

Assumption 2.5 (Smoothness) *$\beta(x)$ is continuous and differentiable in $\mathcal{X}$.*

Under this smoothness restriction we have the following Theorem.

Theorem 2.1 (Identification) *If X_t is continuously-valued and Assumptions 2.1 to 2.5 hold, then β is identified by*

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

with $\beta(x)$ given by (4) 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 2.5 as rather weak. In any case estimation would be impossible without it.

2.1 Aggregate time effects

Although the APE is only partially identified when X_t is discrete and 'just-identified' when X_t is continuous, our CRC model nevertheless has testable implications. In particular the CRC outcome response, marginal stationarity and the absence of time effects imply that:

$$\mathbb{E}\left[\Delta Y | X = x\right] = \mathbb{E}\left[\Delta Y | X = x'\right] = 0,$$

where x and x' denote two different types of 'stayers':

$$\left\{x, x' : x, x' \in \mathcal{X}, \quad x_1 = x_2, \quad x_1' = x_2', \quad x_1 \neq x_1'\right\}.$$

Outcome changes for stayers are driven solely by changes in $a_t(A, U_t)$ and $b_t(A, U_t)$ over time. However, since marginal stationarity and the absence of time effects implies constancy of the conditional means of $a_t(A, U_t)$ and $b_t(A, U_t)$, our model implies that, on average, outcomes do not change across periods for stayers. Since, when X_t is continuous, there may be many types of stayers, corresponding to different values of x_2 (with $x_1 = x_2$), our set-up therefore generates many testable restrictions.

We can use these extra model restrictions to relax Assumption 2.4 and hence incorporate aggregate time effects into our model in a fairly flexible way.

Assumption 2.6 (Conditional Common Average Trends)

$$\begin{aligned}\mathbb{E}\left[a_2(A, U_t) - a_1(A, U_t) | X\right] &= \delta_a(X_2) \\ \mathbb{E}\left[b_2(A, U_t) - b_1(A, U_t) | X\right] &= \delta_b(X_2).\end{aligned}$$

Assumption 2.6 allows for heterogeneity in the period two aggregate time shock across units. In particular, the average shock may differ across subpopulations defined in terms of their period two choice. For example, if X_t denotes union membership, then Assumption 2.6 allows for the period two shock to affect mean earnings in the union and non-union sectors differently.

Let x and x' respectively denote a mover and stayer such that $x_2' = x_2$ (i.e., the mover and stayer have the same period 2 regressor values). Under Assumption 2.6 we have

$$\mathbb{E}\left[Y_2 | X = x'\right] - \mathbb{E}\left[Y_1 | X = x'\right] = \delta_a(x_2) + \delta_a(x_2)\, x_2,$$

and also

$$\mathbb{E}\left[Y_2|\, X=x\right]-\mathbb{E}\left[Y_1|\, X=x\right]=\delta_a\left(x_2\right)+\delta_b\left(x_2\right)x_2+\beta\left(x\right)\left(x_2-x_1\right).$$

We therefore have

$$\beta\left(x\right)=\frac{\mathbb{E}\left[Y_2|\, X=x\right]-\mathbb{E}\left[Y_1|\, X=x\right]-\left\{\mathbb{E}\left[Y_2|\, X=x'\right]-\mathbb{E}\left[Y_1|\, X=x'\right]\right\}}{x_2-x_1}. \tag{7}$$

We may adapt expression (6) to get $\beta\left(x\right)$ for stayers. With $\beta\left(x\right)$ identified, identification of the APE follows directly. Note that $\delta_a\left(x_2\right)$ and $\delta_b\left(x_2\right)$ are not separately identified without further restrictions. In the absence of such restrictions a convenient, and interpretable, normalization is to assume that $\delta_b\left(x_2\right)=0$ for all $x_2\in\mathcal{X}_2$.

Assumption 2.6 is a generalization of the deterministic 'common trends' assumption routinely made in program evaluation studies (Heckman and Robb, 1985; Meyer, 1995; Angrist and Krueger, 1999). In that literature Assumption 2.6 is invoked with the additional requirements that $\delta_a\left(X_2\right)\equiv\delta_a$ is constant in X_2 and $\delta_b\left(X_2\right)=0$. This corresponds to an 'unconditional' common average trends assumption (cf., Heckman, Ichimura, Smith and Todd (1998) for an extensive discussion of a related point).

For estimation purposes it is convenient to assume a parametric forms for $\delta_a\left(X_2\right)$ and $\delta_b\left(X_2\right)$. A simple specification assumes that both $\delta_a\left(X_2\right)$ and $\delta_b\left(X_2\right)$ are constant in X_2. Such a model allows for a fairly flexible pattern of heterogeneous macroeconomic shocks over time, while at the same time remaining easy to interpret and, importantly, easy to estimate. At the same time it provides a set of testable restrictions which may be used to judge model adequacy. Namely that for any two stayers with $X=x'$ and $X=x''$ we have

$$\frac{\mathbb{E}\left[\Delta Y|\, X=x''\right]-\mathbb{E}\left[\Delta Y|\, X=x'\right]}{x_2''-x_2'}=\delta_b.$$

We work with this specification in next subsection and with the even simpler case where $\delta_a\left(X_2\right)=\delta_a$ and $\delta_b\left(X_2\right)=0$ in our initial discussion of estimation. We then return to more general models for aggregate time effects in Section 4 below.

2.2 Relationship to linear 'fixed effects' (FE) and semiparametric 'random effects' analysis

Before discussing estimation we connect the identification results presented above to the textbook within-group regression estimator and the recently proposed semiparametric correlated random effects methodology of Altonji and Matzkin (2005).

2.2.1 Relationship to first differences estimator

Our model can be used to provide a representation of the probability limit of the textbook FE-OLS estimator under CRC misspecification.[11] Assume that the researcher posits a model of

$$Y_t = \delta_t + \beta X_t + A + U_t, \qquad \mathbb{E}\left[U_t \middle| A, X\right] = 0, \quad t = 1, 2 \tag{8}$$

when in fact the true model is as described by Assumptions 2.1, 2.2, 2.3, 2.5 and 2.6 with $\delta_a(X_2) = \delta_a$ and $\delta_b(X_2) = \delta_b$.

In the $T = 2$ case the linear FE estimator has a probability limit equal to the coefficient, b^{FE}, on ΔX in the (mean squared error minimizing) linear predictor of ΔY given ΔX. It is straightforward to show that

$$b^{FE} = \beta + \delta_b \left\{1 - \frac{\mathbb{C}(X_1, X_2)}{\mathbb{V}(\Delta X)}\right\} + \mathbb{E}\left[\omega(\Delta X)(\beta(X) - \beta)\right], \qquad \omega(\Delta X) = \frac{\Delta X(\Delta X - \mathbb{E}[\Delta X])}{\mathbb{V}(\Delta X)}. \tag{9}$$

The first term in (9) reflects the failure of the textbook model to account for aggregate slope drift, while the second is due to its failure to account for slope heterogeneity. This second term is similar to the local average treatment effect (LATE) representation of the Wald-IV estimator's probability limit (Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996; Imbens, 2007). If slope drift is not a concern (i.e., $\delta_b = 0$), we can view b^{FE} as a movers *weighted* average partial effect since $\mathbb{E}[\omega(\Delta X)] = 1$ and $\omega(0) = 0$. An important difference between (9) and the LATE is that 'movers', unlike 'compliers', can be directly identified from the data. Consequently the weights in (9) are estimable.

To get a sense of whether b^{FE} is likely to be interpretable it is helpful to consider some stylized examples. For simplicity we assume, for the remainder of this subsection, the absence of slope drift (i.e., that $\delta_b = 0$). If X_1 and X_2 are independent and identically distributed normal random variables, then $\omega(\Delta X)$ will be a χ^2_1 random variable and b^{FE} will be 'dominated' by those few units with very large values of ΔX. This suggests that b^{FE} will be more representative of the partial effect of those units who change their choice of X_t dramatically across periods.

The binary X_t case is also informative. Let π_{ij} denote the probability that $X_1 = i$ and $X_2 = j$ (with $i, j \in \{0, 1\}$), we can show that

$$b^{FE} = \omega(-1)\beta(0,1) + (1 - \omega(-1))\beta(1,0), \qquad \omega(-1) = \frac{\pi_{01}(1 - \pi_{01} + \pi_{10})}{\pi_{01}(1 - \pi_{01}) + \pi_{10}(1 - \pi_{10}) + 2\pi_{01}\pi_{10}}$$

which is a weighted average of the average partial effect of those units who 'move' from $X_1 = 0$ to $X_2 = 1$ ('joiners') and those who move from $X_1 = 1$ to $X_2 = 0$ ('leavers'). If $\pi_{10} = \pi_{01}$ such that $\mathbb{E}[\Delta X] = 0$, then $b^{FE} = \beta^M$, however, in general the two estimands will differ (this equality also holds when (8) does not include time-specific intercepts).

[11] Chernozhukov, Fernández-Val, Hahn and Newey (2008, Theorem 1) provide a representation result for the fixed effects estimator when the true model is nonlinear and $T > 2$. Their result assumes the absence of any aggregate time effects.

The linear FE estimator is especially interpretable in the 'classical' difference-in-differences (DID) set-up. In that setting there are two sets of regions. In both sets of regions the program is unavailable in period one. In treatment regions it becomes available in period two, while in control regions it remains unavailable. In that case $\pi_{10} = 0$ and it is easy to see that

$$b^{FE} = \beta(0, 1),$$

which also equals the average treatment effect on the treated (ATT). This result shows that, under CRC misspecification, the standard difference-in-differences estimator, while inconsistent for the average partial effect (APE) – the average treatment effect (ATE) is this context – nevertheless has an interpretable probability limit.

Wooldridge (2005b), who maintains the CRC structure as we do, imposes the additional restriction (in our notation) that $\beta(x) = \mathbb{E}[b(A, U_t)]$ for $x_1 \neq x_2$ (cf., Equation (14) on p. 387).[12] In that case equivalency of the FE probability limit and the APE follows directly by the property that $\mathbb{E}[\omega(\Delta X)] = 1$. Chamberlain (1982, p. 11) makes a similar point. He notes, again in our notation, that if $\mathbb{C}(b(A, U_1), X_1) = \mathbb{C}(b(A, U_2), X_2)$, then $\mathbb{E}^*[\beta(X)|\, \Delta X] = \mathbb{E}[\beta(X)]$ so that $\mathbb{E}[\beta(X)|\, \Delta X] = \mathbb{E}[\beta(X)]$. Iterated expectations applied to (9) then gives the equality $b^{FE} = \mathbb{E}[\beta(X)]$.

While, covariance stationarity of the random slopes may be plausible in some settings, it will strain credibility in others. Consider a government which allocates a certain program across regions. Assume that initially, in period 1, the program is regressively targeted in the sense that it is placed in those regions where returns, $b(A, U_1)$, are low, while in period 2 targeting takes an opposite, progressive form. In that case $\mathbb{C}(b(A, U_1), X_1) < 0 < \mathbb{C}(b(A, U_2), X_2)$ and $b^{FE} \neq \mathbb{E}[\beta(X)]$. This example may be of more than intellectual interest: policy 'experiments' are often associated with changes of government or legislation that involves alterations of the implicit targeting rule (e.g., Duflo, 2001). However, in other cases, covariance stationarity may be reasonable. For example, the pattern of selection into unions is plausibly stable across two adjacent years with similar macroeconomic conditions (as in Card, 1996). In any case, our approach does not require these types of restrictions.

2.2.2 Relationship to semiparametric correlated random effects methods

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

$$Y_t = m_t(X_t, A, U_t) \tag{10}$$

and the following exchangeability assumption:

Assumption 2.7 (Exchangeability) *(i)*

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

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

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 2.7, unlike Assumption 2.2 above, *does* restrict the conditional distribution of A given X. Under Assumption 2.7 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.[13] Because Assumption 2.7 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.[14]

Following their second suggestion, the imposition of our CRC structure on (10) and Assumption 2.7 implies that

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

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 long 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 in 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 2.7 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 2.7.

A natural combination of the two assumptions is:

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

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

Assumption 2.8 (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 2.8 $\beta(x)$ is overidentified since, where for simplicity we also maintain Assumption 2.4 (but this is not essential),

$$\beta(x) = \frac{\mathbb{E}[Y_2|\, X = x] - \mathbb{E}[Y_1|\, X = x]}{x_2 - x_1} = \frac{\mathbb{E}[Y_2|\, X' = x'] - \mathbb{E}[Y_1|\, X' = x']}{x'_2 - x'_1},$$

when x' is a permutation of x.

3 Estimation

In this section we discuss estimation of the movers average partial effect, β^M, when the regressors are discretely-valued, and the average partial effect, β, with continuously-valued regressors. To keep the exposition simple we work with the aggregate time effects specification where $\delta_a(X_2) = \delta_a$ and $\delta_b(X_2) = 0$

3.1 Discrete regressor case

We begin with the discrete regressor case, as it straightforward, and foreshadows our estimation approach for continuous regressors. Under our assumptions we can identify the common trend by the average change in Y_t across the two periods among the subpopulation of 'stayers'. That is

$$\delta_a \equiv \mathbb{E}[\Delta Y|\, \Delta X = 0] = \frac{\mathbb{E}[\mathbf{1}(\Delta X = 0)\, \Delta Y]}{\mathbb{E}[\mathbf{1}(\Delta X = 0)]}.$$

We, of course, require that $\mathbb{E}[\mathbf{1}(\Delta X = 0)] = \Pr(\Delta X = 0)$ is greater than zero: *it is the presence of stayers which identifies δ_a*. Now consider the subpopulation of movers, we have

$$\mathbb{E}[\Delta Y|\, X = x] = \delta_a + \beta(x)\, \Delta x,$$

and hence, with δ_a identified, we may write

$$\beta^M \equiv \frac{\mathbb{E}\left[\mathbf{1}(\Delta X \neq 0)\, \frac{\mathbb{E}[\Delta Y|X] - \delta_a}{\Delta X}\right]}{\mathbb{E}[\mathbf{1}(\Delta X \neq 0)]} = \frac{\mathbb{E}\left[\mathbf{1}(\Delta X \neq 0)\, \frac{\Delta Y - \delta_a}{\Delta X}\right]}{\mathbb{E}[\mathbf{1}(\Delta X \neq 0)]}.$$

Let $\theta = \left(\delta_a, \beta^M\right)'$, the above expressions generate the following 2×1 vector of moment restrictions $\mathbb{E}\left[\psi\left(Z, \theta_0\right)\right] = 0$, with

$$\psi\left(Z,\theta\right) = \left(\begin{array}{c} \mathbf{1}\left(\Delta X = 0\right)\left(\Delta Y - \delta_a\right) \\ \frac{\mathbf{1}\left(\Delta X \neq 0\right)}{\Delta X}\left(\Delta Y - \delta_a - \beta^M \Delta X\right) \end{array} \right).$$

The GMM estimate $\widehat{\beta}^M$ is very easy to compute, being the coefficient on ΔX in the linear instrumental variables fit of ΔY on a constant and ΔX with $\mathbf{1}\left(\Delta X = 0\right)$ and $\frac{\mathbf{1}\left(\Delta X \neq 0\right)}{\Delta X}$ serving as excluded instruments (this follows since $\mathbf{1}\left(\Delta X = 0\right)\left(\Delta Y - \delta_a - \beta^M \Delta X\right) = \mathbf{1}\left(\Delta X = 0\right)\left(\Delta Y - \delta_a\right)$). Conventional 'robust' standard errors reported by most software packages will be asymptotically valid.

Since it foreshadows portions of our results for the continuous X_t case we present a closed-form expression for the asymptotic sampling variance of $\widehat{\beta}^M$. Let $\Gamma_0 = \mathbb{E}\left[\partial \psi\left(Z, \theta_0\right) / \partial \theta'\right]$ and $\Omega_0 = \mathbb{E}\left[\psi\left(Z, \theta_0\right)\psi\left(Z, \theta_0\right)'\right]$ and further define

$$\pi_0 = \Pr\left(\Delta X = 0\right), \qquad \sigma_0^2 = \mathbb{V}\left(Y \middle| \Delta X = 0\right)$$
$$\xi = \mathbb{E}\left[\frac{1}{\Delta X}\middle| \Delta X \neq 0\right], \qquad \kappa = \mathbb{E}\left[\mathbb{V}\left[\frac{\Delta Y}{\Delta X}\middle| X\right]\middle| \Delta X \neq 0\right] + \mathbb{V}\left(\beta\left(X\right)\middle| \Delta X \neq 0\right).$$

We have

$$\Gamma_0 = -\left(\begin{array}{cc} \pi_0 & 0 \\ \left(1-\pi_0\right)\xi & \left(1-\phi_0\right) \end{array} \right), \qquad \Omega_0 = \left(\begin{array}{cc} \pi_0 \sigma_0^2 & 0 \\ 0 & \left(1-\pi_0\right)\kappa \end{array} \right),$$

and hence, by standard results for GMM (e.g., Newey and McFadden, 1994), an asymptotic sampling distribution for $\widehat{\theta}$ of

$$\sqrt{N}\left(\begin{array}{c} \widehat{\delta}_a - \delta_a \\ \widehat{\beta}^M - \beta^M \end{array} \right) \xrightarrow{d} \mathcal{N}\left(\underline{0}, \left(\begin{array}{cc} \frac{\sigma_0^2}{\pi_0} & -\frac{\sigma_0^2}{\pi_0}\xi \\ -\frac{\sigma_0^2}{\pi_0}\xi & \kappa + \frac{\sigma_0^2}{\pi_0}\xi^2 \end{array} \right)\right). \tag{11}$$

Two features of (11) reappear in the continuous case. First, the precision on the estimated common trends depends on the variance of the ΔY amongst stayers, σ_0^2, as well as their population frequency, π_0. As the frequency of stayers increases, so does our ability to precisely estimate aggregate time effects. Second, the asymptotic sampling variance of $\widehat{\beta}^M$ depends on the distribution of the regressors through the term ξ^2. If $\xi \neq 0$, as would be the case if there is positive drift in X_t over time, then the variance of $\widehat{\beta}^M$ will have a component which depends on the precision with which we can estimate δ_a. If instead $\xi = 0$, the asymptotic properties of $\widehat{\beta}^M$ do not depend on those of $\widehat{\delta}_a$. In that case we can estimate $\widehat{\beta}^M$ as precisely as we could if we somehow knew δ_a *a priori.*

3.2 Continuous regressor case

When X_t is continuously valued then, under smoothness conditions, the APE is identified. However continuity of X_t raises technical issues, the resolution of which require the use of nonparametric methods. As a result our estimates of β generally converge at the one-dimensional nonparametric rate. To highlight the issues involved we first discuss estimation of β in the absence of time effects, followed by a discussion of time effects and finally how local-linear methods can be used when the distribution X_t has mass points at a finite number of values.

3.2.1 No time effect

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 ($\delta_a(X_2) = \delta_b(X_2) = 0$), then Theorem 2.1 implies that the average partial effect β is identified by

$$\beta = \mathbb{E}\left[\frac{\mathbb{E}[\Delta Y|X]}{\Delta X}\right] = \mathbb{E}\left[\frac{\mathbb{E}[\Delta Y|X]}{\Delta X}\,|\Delta X \neq 0\right].$$

Given the second equality a natural estimator of the APE, β, would be that proposed for the discrete case above, that is,

$$\widetilde{\beta} = \frac{\sum_{i=1}^{N}\mathbf{1}(\Delta X_i \neq 0)\left(\frac{\Delta Y_i}{\Delta X_i}\right)}{\sum_{i=1}^{N}\mathbf{1}(\Delta X_i \neq 0)} = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta Y_i}{\Delta X_i}.$$

This estimator was informally suggested by Mundlak (1961, p.45); as Chamberlain (1980b, 1982) notes, it will be strongly consistent if $\mathbb{E}[|\Delta Y/\Delta X|] < \infty$ by the strong law of large numbers. However, if ΔX has a positive, continuous density at zero – and if $\mathbb{E}[|\Delta Y| \mid \Delta X = d]$ does not vanish at $d = 0$ – then $\widetilde{\beta}$ will be inconsistent in general, since $\Delta Y/\Delta X$ will not have finite expectation (unlike $\beta(X) = \mathbb{E}[\Delta Y|X]/\Delta X$ whose expectation exists by assumption). For example, if (Y_t, X_t) is independently and identically distributed according to the bivariate normal distribution then $\Delta Y/\Delta X$ will be distributed according to the Cauchy distribution.

To ensure quadratic-mean convergence, we consider instead a 'trimmed' estimator of the form

$$\widehat{\beta}(h_N) \equiv \frac{\sum_{i=1}^{N}\mathbf{1}(|\Delta X_i| > h_N)\left(\frac{\Delta Y_i}{\Delta X_i}\right)}{\sum_{i=1}^{N}\mathbf{1}(|\Delta X_i| > h_N)}, \tag{12}$$

where h_N is a deterministic bandwidth sequence tending to zero as N tends to infinity.[15]

The estimator $\widehat{\beta}(h_N)$ – which is consistent for β^M when X has discrete 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

$$\mathbb{V}(\widehat{\beta}) = O\left(\frac{1}{Nh_N}\right) \gg O\left(\frac{1}{N}\right),$$

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

$$\mathbb{E}\left[\mathbf{1}(|\Delta X| > h_N)\left(\frac{\Delta Y}{\Delta X}\right) - \beta(X)\right] = \mathbb{E}\left[\mathbf{1}(|\Delta X| \leq h_N)\beta(X)\right] = O(h_N^p)$$

[15] An alternative consistent estimator would replace the denominator by the sample size N.

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

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

which yields

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

While the bandwidth sequence h_N^* achieves the fastest rate of convergence for this estimator, the corresponding asymptotic normal distribution for $\widehat{\beta}(h_N^*)$ will be centered at a bias term involving the derivative of $\mathbb{E}[\beta(X)|\Delta X = d]$ at $d = 0$. The estimator $\widehat{\beta}$ 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 $\widehat{\beta}$,

$$\sqrt{Nh_N}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, 2\phi_0\sigma_0^2),$$

where

$$\phi_0 \equiv \lim_{h \downarrow 0} \frac{\Pr\{|\Delta X| \leq h\}}{2h}$$

is the density of ΔX at zero and

$$\sigma_0^2 \equiv \mathbb{V}\left(\Delta Y | \, \Delta X = 0\right) = \lim_{h \downarrow 0} \mathbb{V}\left(\Delta Y | - h < \Delta X < h\right).$$

Assuming $p = 2$, the asymptotic distribution of $\widehat{\beta}$ is similar to the asymptotic distribution of a (uniform) kernel regression estimator of $\mathbb{E}[\Delta Y|\Delta X = 0]$, except that the variance of the latter varies inversely, not directly, with the density ϕ_0.

3.2.2 Aggregate time effect

When aggregate time effects are present, and the 'common trends' condition (Assumption 2.6) holds with $\delta_a(X_2) = \delta_a$ and $\delta_b(X_2) = 0$, then (7) implies that the average partial effect β is identified by

$$\beta = \mathbb{E}\left[\frac{\mathbb{E}\left[\Delta Y | \, X\right] - \delta_a}{\Delta X}\right] = \mathbb{E}\left[\frac{\mathbb{E}\left[\Delta Y | \, X\right] - \delta_a}{\Delta X}\middle| \Delta X \neq 0\right],$$

recalling that $\delta_a \equiv \mathbb{E}[\Delta Y|\Delta X = 0]$. If δ_a were known, a straightforward modification of the estimator proposed in the preceding section would be

$$\widehat{\beta}_I(h_N) = \frac{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)\left(\frac{\Delta Y_i - \delta_a}{\Delta X_i}\right)}{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)},$$

which would inherit the large sample properties of $\widehat{\beta}(h_N)$ above.

When δ_a is unknown, a natural counterpart to the infeasible estimator $\widehat{\beta}_I(h_N)$ replaces δ_a with the uniform kernel estimator,

$$\widehat{\delta}_a(h_N) \equiv \frac{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| \leq h_N)\Delta Y_i}{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| \leq h_N)}, \tag{13}$$

whose asymptotic properties are well-known when ΔX is continuously distributed. Under standard regularity conditions a normalized version of $\widehat{\delta}_a$ has the asymptotic distribution,

$$\sqrt{Nh_N}(\widehat{\delta}_a - \delta_a) \xrightarrow{d} \mathcal{N}(0, \sigma_0^2/2\phi_0),$$

where ϕ_0 and σ_0^2 are defined above. Furthermore, $\widehat{\beta}_I$ and $\widehat{\delta}_a$ are asymptotically independent, as the product of their influence functions is zero by construction.

Given this estimator of the common trend δ_a, a feasible estimator of the APE β is

$$\widehat{\beta}_F(h_N) = \frac{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)\left(\frac{\Delta Y_i - \widehat{\delta}_a(h_N)}{\Delta X_i}\right)}{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)}. \tag{14}$$

Though simple in appearance, derivation of the large-sample properties of $\widehat{\beta}_F$ is difficult, as its rate of convergence depends in a delicate way on the distribution of the regressors X. Some of these issues were foreshadowed by our discussion of the discrete case above. Writing the normalized version of $\widehat{\beta}_F$ in terms of its infeasible counterpart $\widehat{\beta}_I$ yields

$$\sqrt{Nh_N}(\widehat{\beta}_F - \beta) = \sqrt{Nh_N}(\widehat{\beta}_I - \beta) - \sqrt{Nh_N}(\widehat{\delta}_a - \delta_a) \times \left[\frac{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)\left(\frac{1}{\Delta X_i}\right)}{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)}\right].$$

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

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

will crucially depend upon the behavior of

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

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 $\widehat{\beta}_F$ and its infeasible counterpart $\widehat{\beta}_I$. Alternatively, if there is constant positive drift in the distribution of regressors, so that $\tau(0) > 0$, then the third term $\widehat{\xi}$ 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 $\widehat{\beta}_F$ will be dominated by the asymptotic distribution of $\widehat{\delta}_a$, the estimator 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 $\widehat{\beta}_F$ can be constructed if consistent estimators of the density ϕ_0 and conditional variance σ_0^2 terms appearing in the asymptotic variances of $\widehat{\beta}_I$ and $\widehat{\delta}_a$ can be constructed. Under standard regularity conditions, the kernel estimators

$$\widehat{\phi} \equiv \frac{1}{2Nh_N} \sum_{i=1}^{N} \mathbf{1}(|\Delta X_i| \leq h_N), \qquad \widehat{\sigma}^2 \equiv \frac{\sum_{i=1}^{N} \mathbf{1}(|\Delta X_i| \leq h_N)\,(\Delta Y_i)^2}{\sum_{i=1}^{N} \mathbf{1}(|\Delta X_i| \leq h_N)} - \widehat{\delta}_a^2$$

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

$$\widehat{AVar}(\widehat{\beta}_F) = \frac{\widehat{\sigma}^2}{Nh_N} \left(2\widehat{\phi} + \frac{\widehat{\xi}^2}{2\widehat{\phi}} \right),$$

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

3.2.3 Mixed discrete-continuous regressors

In some applications the distribution of the regressors (X_1, X_2) may have mass points at a finite set of values, while being 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 β will generally differ from its 'movers' counterpart β^M, due to the nonzero probability that $\Delta X = 0$; while this mass point simplifies estimation of a nonzero common trend component δ_a (and the conditional variance of ΔY given $\Delta X = 0$), it complicates estimation of the APE. This is because β typically differs from β^M, which is the implicit estimand of (12) and (14) above, when 'stayers' are a non-negligible portion of the population.

When $\pi_0 \equiv \Pr(\Delta X = 0) > 0$, the estimator

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

used for the discrete X_t case discussed above, is clearly $\sqrt{N}$-consistent and asymptotically normal estimator for δ_a, as would be the (asymptotically equivalent) estimator $\widehat{\delta}_a$ defined in the previous subsection. Using the decomposition for the feasible estimator $\widehat{\beta}_F$ of $\beta^M \equiv \mathbb{E}[\beta(X)|\Delta X \neq 0]$ in the previous section, it follows that

$$\begin{aligned}\sqrt{Nh_N}(\widehat{\beta}_F - \beta^M) &= \sqrt{Nh_N}(\widehat{\beta}_F - \beta^M) + O_p(\sqrt{h_N}) \cdot O_p(\log(h_n^{-1})) \\ &= \sqrt{Nh_N}(\widehat{\beta}_F - \beta^M) + o_p(1),\end{aligned}$$

so that preliminary estimation of the common trend component δ_a will not affect the asymptotic distribution of the feasible estimator $\widehat{\beta}_F$. If a consistent estimator of the stayers effect

$$\beta^S \equiv \mathbb{E}[\beta(X)|\ \Delta X\ = 0]$$

can be constructed, a corresponding consistent estimator of the APE $\beta = \pi_0\beta^S + (1-\pi_0)\beta^M$ would be

$$\widehat{\beta} \equiv \widehat{\pi}\widehat{\beta}^S + (1-\widehat{\pi})\widehat{\beta}_F,$$

where $\widehat{\pi} \equiv \sum_{\iota=1}^{N} \mathbf{1}(|\Delta X| \leq h_N)/N$ is a $\sqrt{N}$-consistent estimator for π_0.

Defining

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

the results of Section 2 above imply that

$$\beta^S = \lim_{h\downarrow 0} \frac{\nu(h) - \nu(0)}{h};$$

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

$$\begin{pmatrix} \overline{\delta}_a \\ \widehat{\beta}^S \end{pmatrix} = \arg\min_{d_a, b^S} \sum_{i=1}^{N} \mathbf{1}(|\Delta X_i| \leq h_N) \cdot (\Delta Y_i - d_a - b^S \Delta X_i)^2, \tag{15}$$

with the intercept $\overline{\delta}_a$ being an alternative ($\sqrt{N}$-)consistent estimator of the common trend δ_a. 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 $\widehat{\beta} \equiv \widehat{\pi}\widehat{\beta}^S + (1-\widehat{\pi})\widehat{\beta}_F$ of the APE will be the same as for $\widehat{\beta}^S$, and the asymptotic distribution of the latter would dominate the asymptotic distribution of $\widehat{\beta}$ in this setting.

3.2.4 Computation

For estimation of, and inference on, the APE we propose using a simple instrumental variables (IV) procedure. Consider the instrumental variables fit associated with the linear regression of ΔY on a constant and the interactions $\mathbf{1}(|\Delta X| > h_N) \cdot \Delta X$ and $\mathbf{1}(|\Delta X| \leq h_N) \cdot \Delta X$ with $\mathbf{1}(|\Delta X| \leq h_N)$, $\mathbf{1}(|\Delta X| \leq h_N) \cdot \Delta X$ and $\frac{\mathbf{1}(|\Delta X| > h_N)}{\Delta X}$ serving as excluded instruments. The coefficients on the first two regressors will equal $\overline{\delta}_a$ and $\widehat{\beta}^S$ as defined by (15) above, while the coefficient on the last regressor will equal $\widehat{\beta}_F$ (as given in (14) with $\overline{\delta}_a$ replacing $\widehat{\delta}_a$). The robust standard errors reported by most statistical packages will be asymptotically valid.[16]

If the mixed discrete-continuous case discussed above is of relevance, then we may combine these 'IV moment conditions' together with the 'moment' $\mathbf{1}(|\Delta X| \leq h_N) - \pi$ to form a single quasi-GMM problem. We can then estimate of the average partial effect by $\widehat{\beta} \equiv \widehat{\pi}\widehat{\beta}^S + (1-\widehat{\pi})\widehat{\beta}_F$. A combination of the conventional GMM covariance matrix and the textbook delta method may be used to form standard errors for $\widehat{\beta}$.

4 Multiple regressors and time periods

In this section we extend our basic model to permit multiple regressors and panels of arbitrary length. Formally we analyze the following correlated random coefficients model:

$$Y_t = \mathbf{W}_t'\mathbf{d}\left(A, U_t\right) + \mathbf{X}_t'\mathbf{b}\left(A, U_t\right), \quad t = 1, ..., T,$$

where $\mathbf{W}_t$ and $\mathbf{X}_t$ are $q \times 1$ and $p \times 1$ vectors of observable regressors and $\mathbf{d}\left(A, U_t\right)$ and $\mathbf{b}\left(A, U_t\right)$ corresponding random coefficients (all with bounded moments).

Our marginal stationarity restriction is

$$U_t | \, \mathbf{W}, \mathbf{X}, A \sim U_s | \, \mathbf{W}, \mathbf{X}, A,$$

for $s \neq t$, $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_T)'$ and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)'$. This implies that

$$\mathbb{E}[\mathbf{d}\left(A, U_t\right) | \mathbf{W}, \mathbf{X}] = \boldsymbol{\delta}\left(\mathbf{W}, \mathbf{X}\right)$$

and

$$\mathbb{E}[\mathbf{b}\left(A, U_t\right) | \mathbf{W}, \mathbf{X}] = \boldsymbol{\beta}\left(\mathbf{W}, \mathbf{X}\right).$$

To complete the model we make the additional restrictions that

$$\boldsymbol{\delta}\left(\mathbf{W}, \mathbf{X}\right) \equiv \boldsymbol{\delta}, \quad \boldsymbol{\beta}\left(\mathbf{W}, \mathbf{X}\right) = \boldsymbol{\beta}\left(\mathbf{X}\right).$$

In this model $\mathbf{W}$ is a $T \times q$ matrix of aggregate 'time shifters'. Typically we think of these regressors as varying deterministically with t, and hence the coefficients $\mathbf{d}\left(A, U_t\right)$ as capturing time- and individual-specific trends. The $T \times p$ matrix of regressors $\mathbf{X}$ includes the choice/policy variables of primary interest.

The two period model considered in the preceding sections is contained within the above family

[16] Note these standard errors will implicitly include estimates of asymptotically negligible terms. However, this may improve small sample coverage of the resulting confidence intervals (cf., Newey 1994b).

with $T = p = 2$ and $q = 1$. The matrix of time shifters and its corresponding coefficient vector parameterize the common intercept shift across periods:

$$\mathbf{W} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{\delta} = \delta_a,$$

while the choice variable and the conditional means of the random coefficients are given by

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \end{pmatrix}, \quad \boldsymbol{\beta}(\mathbf{X}) = \begin{pmatrix} \alpha(X) \\ \beta(X) \end{pmatrix}.$$

As before, the parameters of interest are $\boldsymbol{\delta} \equiv \mathbb{E}[\mathbf{d}(A, U_t)]$ and $\beta \equiv \mathbb{E}[\mathbf{b}(A, U_t)]$.

The above model is a special case of the CRC model proposed and analyzed by Chamberlain (1992a), who worked with a more general setup where regressors and trend coefficients were permitted to vary parametrically (i.e., $\mathbf{W} = \mathbf{W}(\boldsymbol{\theta})$, $\mathbf{X} = \mathbf{X}(\boldsymbol{\theta})$, and $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\theta})$). Identification of $\boldsymbol{\delta}$ and β in the overidentified setup $T > p$ was considered in detail by Chamberlain (1992a), we begin with 'just identified' case $T = p$, which he did not consider, and return to the overidentified case subsequently.

4.1 Just identification

Writing $\mathbf{Y} = (Y_1, ..., Y_T)'$ we have

$$\mathbb{E}[\mathbf{Y}|\mathbf{W}, \mathbf{X}] = \mathbf{W}\boldsymbol{\delta} + \mathbf{X}\boldsymbol{\beta}(\mathbf{X}). \tag{16}$$

Define $\tilde{X}$ to be the (scalar) determinant of the matrix of regressors,

$$\tilde{X} = \det(\mathbf{X}),$$

and $\mathbf{X}^*$ to be the *adjoint* (or *adjunct*) matrix to $\mathbf{X}$, i.e., the transpose of the matrix of cofactors of $\mathbf{X}$,

$$\mathbf{X}^* \equiv adj(\mathbf{X}),$$

so that, $\mathbf{X}^*\mathbf{X} = \tilde{X} \cdot \mathbf{I}$, and, when $\tilde{X} \neq 0$, $\mathbf{X}^{-1} = (1/\tilde{X}) \cdot \mathbf{X}^*$ (recall that with $T = p$ that $\mathbf{X}$ is a square matrix). Premultiplication of the vector of conditional means of Y_t by the adjoint matrix $\mathbf{X}^*$ thus yields

$$\mathbb{E}[\mathbf{X}^*\mathbf{Y}|\mathbf{W}, \mathbf{X}] = \mathbf{X}^*\mathbf{W}\boldsymbol{\delta} + \tilde{X} \cdot \boldsymbol{\beta}(\mathbf{X}),$$

which implies that

$$\mathbb{E}[\mathbf{X}^*\mathbf{Y}|\mathbf{X}, \mathbf{W}, \tilde{X} = 0] = \mathbf{X}^*\mathbf{W}\boldsymbol{\delta},$$

assuming $\mathbf{Y}$ and $\mathbf{X}$ have at least $T + 1$ moments finite (ensuring $E[||\mathbf{X}^*\mathbf{Y}||] < \infty$).

Provided the random $(T \times q)$ matrix $\mathbf{X}^*\mathbf{W}$ has q-dimensional support conditional on $\tilde{X} = 0$, the coefficient vector $\boldsymbol{\delta}$ is identified by a population regression of $\mathbf{X}^*\mathbf{Y}$ on $\mathbf{X}^*\mathbf{W}$ conditional on $\tilde{X} \equiv \det(\mathbf{X}) = 0$. By analogy with the estimation results for the scalar case presented above, a consistent estimator of $\boldsymbol{\delta}$ can be constructed using a weighted least-squares regression of $\mathbf{X}_i^*\mathbf{Y}_i$ on $\mathbf{X}_i^*\mathbf{W}_i$ across all observations $i = 1, ..., N$, with weights equal to $\mathbf{1}(|\tilde{X}_i| \leq h_N)$. Thus, estimation of $\boldsymbol{\delta}$ still involves a one-dimensional nonparametric regression problem in the (scalar) conditioning variable $\tilde{X}_i$.

In the $T = 2$ case considered in the preceding sections we have

$$\tilde{X} = \det \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \end{pmatrix} = \Delta X,$$

so that

$$\mathbf{X}^*\mathbf{W}\boldsymbol{\delta} = \begin{bmatrix} X_2 & -X_1 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta_a = \begin{pmatrix} -X_1\delta_a \\ \delta_a \end{pmatrix},$$

and

$$\mathbf{X}^*\mathbf{Y} = \begin{pmatrix} X_2Y_1 - X_1Y_2 \\ \Delta Y \end{pmatrix}.$$

When $\tilde{X} = \Delta X = 0$, the two rows of $\mathbf{X}^*\mathbf{Y} - \mathbf{X}^*\mathbf{W}\boldsymbol{\delta}$ are proportional to each other, and either could be used to define a nonparametric estimator of δ_a; in the preceding sections, the second row was used.

Returning to the general case with $T = p \geq 2$, given identification of $\boldsymbol{\delta}$, identification of β^M follows from the equality

$$\mathbb{E}[\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}|\mathbf{W}, \mathbf{X}] = \mathbf{X}\boldsymbol{\beta}(\mathbf{X}).$$

When $\tilde{X} \equiv \det(\mathbf{X}) \neq \mathbf{0}$, premultiplying both sides of this relation by $\mathbf{X}^{-1}$ yields

$$\mathbf{E}[\mathbf{X}^{-1}(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid \mathbf{X} = \mathbf{x}] \equiv \boldsymbol{\beta}(\mathbf{x}),$$

so that, assuming $\Pr(\tilde{X} \neq 0) > 0$

$$\mathbb{E}[\mathbf{X}^{-1}(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid \tilde{X} \neq 0] = \mathbb{E}[\boldsymbol{\beta}(\mathbf{X}) \mid \tilde{X} \neq 0] \equiv \boldsymbol{\beta}^M$$

by iterated expectations.

If $\tilde{X} \neq 0$ with probability one, then the movers average partial effect coincides with the overall or full average partial effect (i.e., $\beta^M = \beta = \mathbb{E}[\mathbf{b}(A, U_t)]$). Heuristically, β^M is identified as an average of a generalized least-squares regression of the detrended conditional mean $\mathbb{E}[\mathbf{Y}|\mathbf{W}, \mathbf{X}] - \mathbf{W}\boldsymbol{\delta}$ on $\mathbf{X}$, averaging over those observations with $\tilde{X} = \det(\mathbf{X}) \neq \mathbf{0}$.

Because the expectation of $\mathbf{X}^{-1}(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta})$ will generally be undefined when $\tilde{X}$ is continuously distributed with positive density near zero, estimation of β^M will involve the same trimmed mean as discussed for the special $T = p = 2$ case above. The extension of the feasible estimator $\widehat{\boldsymbol{\beta}}_F$ to this context is

$$\widehat{\boldsymbol{\beta}}_F = \frac{\sum_{i=1}^N \mathbf{1}(\left|\tilde{X}_i\right| > h_N) \cdot \mathbf{X}_i^{-1}(\mathbf{Y}_i - \mathbf{W}_i\widehat{\boldsymbol{\delta}})}{\sum_{i=1}^N \mathbf{1}(\left|\tilde{X}_i\right| > h_N)}, \tag{17}$$

where $\widehat{\boldsymbol{\delta}}$ is the nonparametric estimator

$$\widehat{\boldsymbol{\delta}} = \left[\sum_{i=1}^{N} \mathbf{1}(\left|\tilde{X}_i\right| \leq h_N)\,(\mathbf{X}_i^*\mathbf{W}_i)'\,(\mathbf{X}_i^*\mathbf{W}_i)\right]^{-1} \times \sum_{i=1}^{N} \mathbf{1}(\left|\tilde{X}_i\right| \leq h_N)\,(\mathbf{X}_i^*\mathbf{W}_i)'\,(\mathbf{X}^*\mathbf{Y})\,. \qquad (18)$$

This estimator will converge in probability to β^M at a one-dimensional nonparametric rate if $h_N \to 0$ at the appropriate rate, provided the term

$$\hat{\boldsymbol{\xi}} \equiv \frac{\sum_{i=1}^{N} 1(\left|\tilde{X}_i\right| > h_N) \cdot \mathbf{X}_i^{-1}\mathbf{W}_i}{\sum_{i=1}^{N} 1(\left|\tilde{X}_i\right| > h_N)}$$

does not diverge too quickly as $N \to \infty$.

In the mixed discrete-continuous case $\Pr(\tilde{X} = 0) > 0$, and estimation of $\hat{\beta}$ requires estimation of

$$\boldsymbol{\beta}^S = \lim_{h \downarrow 0} \frac{\boldsymbol{\nu}(h) - \boldsymbol{\nu}(0)}{h},$$

where $\boldsymbol{\nu}(x) \equiv E[\mathbf{X}^{-1}\mathbf{Y}|\tilde{X} = x]$; the resulting estimator converges at the rate for nonparametric estimation of the derivative of a one-dimensional regression function.

4.2 Overidentification

When $T > p$, the vector of common trend parameters $\boldsymbol{\delta}$ will satisfy some conditional moment restrictions, and, as Chamberlain (1992a) shows, these may suffice for identification and construction of root-N-consistent and asymptotically-normal estimators of $\boldsymbol{\delta}$. In this overidentified setting, for each realized matrix of regressors $\mathbf{X}$ there will be a $T \times (T-p)$ matrix $\mathbf{Z} \equiv \boldsymbol{\zeta}(\mathbf{X})$ of functions of $\mathbf{X}$ for which

$$\mathbf{Z}'\mathbf{X} = \mathbf{0};$$

from the relation (16) above, it follows that

$$\begin{aligned} \mathbf{Z}'\mathbb{E}[\mathbf{Y}|\mathbf{W},\mathbf{X}] &\equiv \mathbf{Z}'\mathbf{W}\boldsymbol{\delta} + \mathbf{Z}'\mathbf{X}\boldsymbol{\beta}(\mathbf{X}) \\ &= \mathbf{Z}'\mathbf{W}\boldsymbol{\delta}, \end{aligned}$$

so that

$$\mathbb{E}[\mathbf{Z}'(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid \mathbf{W},\mathbf{X}] = \mathbf{0},$$

which, depending upon the form of $\mathbf{W}$, will typically serve to identify the trend coefficients $\boldsymbol{\delta}$.

For example, in the $T = 2$ example considered above, suppose the restriction $\alpha(x) \equiv \alpha$ is imposed, so that

$$\boldsymbol{\delta} \equiv (\alpha, \delta_a)', \quad \mathbf{W} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{X} \equiv (X_1, X_2)';$$

then, taking $\mathbf{Z} = (X_2, -X_1)'$, the parameters α and δ_a will satisfy

$$\mathbb{E}[X_2Y_1 - X_1Y_2 - \alpha(X_1 + X_2) - \delta_a X_1 | X_1, X_2] = 0,$$

which implies that α and δ_a will be identified as population least-squares regression coefficients of $X_2Y_1 - X_1Y_2$ on $(X_1 + X_2)$ and X_1, respectively. Alternatively, restricting $\beta(x) = \beta$ but leaving $a(x)$ unrestricted, δ_a and β will be identified by the population regression of ΔY on a constant and ΔX, that is, the population analogue of the usual fixed-effects regression estimator.

Even in the just-identified setting ($T = p$), it may be possible to obtain consistent estimators of $\boldsymbol{\delta}$ that achieve the parametric rate of convergence. If

$$\tilde{\mathbf{W}} \equiv \mathbf{W} - \mathbb{E}[\mathbf{W} \mid \mathbf{X}]$$

has a covariance matrix of full rank, then $\boldsymbol{\delta}$ will be identified by

$$\boldsymbol{\delta} = \mathbb{V}(\tilde{\mathbf{W}})^{-1}\mathbb{C}(\tilde{\mathbf{W}}, \mathbf{Y}).$$

For the special cases considered above, where $\mathbb{V}(\tilde{\mathbf{W}}) = 0$, this is not applicable, but such restrictions may be useful when $\mathbf{W}$ includes regressors which are not deterministic functions of $\mathbf{X}$ even when $T = p$.

Overidentification also makes estimation of β less problematic. As Chamberlain (1992a) shows, defining

$$\widehat{\boldsymbol{\beta}}_i \equiv (\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{W}_i\widehat{\boldsymbol{\delta}})$$

for $\widehat{\boldsymbol{\delta}}$ a root-N-consistent estimator of $\boldsymbol{\delta}$ and $\mathbf{V}_i \equiv \boldsymbol{\nu}(\mathbf{W}_i, \mathbf{X}_i)$ positive definite with probability one, the sample mean of $\widehat{\boldsymbol{\beta}}_i$ will be a root-N-consistent estimator of β when $\mathbf{V}_i = \mathbb{V}(\mathbf{Y}_i - \mathbf{W}_i\boldsymbol{\delta} | \mathbf{X}_i)$ and

$$\mathbb{E}\left[\frac{1}{\det(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})}\right] < \infty. \tag{19}$$

This estimator also attains the semiparametric efficiency bound for estimation of β. Chamberlain (1992a) shows that a feasible version, based upon an efficient estimator of $\boldsymbol{\delta}$ and consistent estimators of $\{\mathbf{V}_i\}_{i=1}^N$, will also be semiparametrically efficient.

As the order of overidentification $T - p$ increases, condition (19) becomes less restrictive even if the components of $\mathbf{X}$ are continuously distributed. For example, consider the $p = 2$ case with $\mathbf{X}_t = (1, X_t)'$ and suppose that $X_t \overset{iid}{\sim} \mathcal{N}(0, 1)$ and $\mathbf{V}_i \equiv \mathbf{I}$; then

$$\det(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i) = \sum_{t=1}^{T}(X_t - \bar{X})^2 \sim \chi^2_{T-1},$$

and (19) will hold as long as $T - 1 > 2$, i.e., $T \geq 4$ here. More generally, as $T - p$ increases, the density of $\det(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)$ should approach zero more rapidly as its argument approaches zero – ensuring (19) holds – provided the continuous components of $\mathbf{X}_i$ are weakly dependent across rows and the matrix $\mathbf{V}_i$ is well-behaved.

Nevertheless, the trimming scheme used to estimate β in the just-identified setting may still be helpful in the overidentified case, even when (19) holds. Defining the (infeasible) trimmed mean

$$\hat{\boldsymbol{\beta}} = \frac{\sum_{i=1}^{N} \mathbf{1}(\det(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i) > h_N) \cdot (\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{W}_i\boldsymbol{\delta})}{\sum_{i=1}^{N} \mathbf{1}(\det(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i) > h_N)},$$

it is straightforward to show this will be asymptotically equivalent to the sample mean of $\hat{\beta}_i$ when $\mathbb{E}[\boldsymbol{\beta}(\mathbf{X})|\det(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i) \leq h]$ is smooth (Lipschitz-continuous) in h, condition (19) holds, and $h_N = o(1/\sqrt{N})$. Since $\hat{\beta}$ will still be consistent for β even when (19) fails, a feasible version of the trimmed mean $\hat{\beta}$ may be better behaved in finite samples if the design matrix $(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)$ is nearly singular for some observations.

5 Empirical application: the demand for calories

5.1 Model and estimation

We assume that the logarithm of total household calorie availability per capita in period t, $\ln(\text{Cal}_t)$, varies according to

$$\ln(\text{Cal}_t) = a_t(A, U_t) + b_t(A, U_t)\ln(\text{Exp}_t) + c_t(A, U_t)\text{Exp}_t^{-1} \tag{20}$$

where Exp_t denote real household expenditure per capita in year t and $a_t(A, U_t)$, $b_t(A, U_t)$, $c_t(A, U_t)$ are random coefficients. The household-by-period-specific elasticity of calorie demand equals

$$\epsilon_t(\text{Exp};\ A, U_t) = b_t(A, U_t) - c_t(A, U_t)\text{Exp}_t^{-1}, \tag{21}$$

which is similar to a heterogenous 'rank three' Engel curve specification (e.g., Lewbel 1991). We use this specification both because it saturates the identifying power of our three period panel and because the large number of very poor households in our sample suggests the need to allow for nonlinearity in the calorie 'Engel' curve. For $b_t(A, U_t) > 0$ and $c_t(A, U_t) < 0$ (21) implies a calorie elasticity which declines with total outlay toward $b_t(A, U_t)$. Strauss and Thomas (1995) and Subramanian and Deaton (1996) discuss the arguments and evidence for an elasticity of calorie demand which declines with total outlay.

Let $X_t = \left(1, \ln(\text{Exp}_t), \text{Exp}_t^{-1}\right)'$ and $Y_t = \ln(\text{Cal}_t)$ with $\mathbf{X}$ and $\mathbf{Y}$ as defined above. Letting $t = 0, 1, 2$ denote the 2000, 2001 and 2002 waves of our panel we allow for common intercept and slope drift of the form

$$\begin{aligned}
\mathbb{E}\left[a_1(A, U_1) - a_0(A, U_0)\right|\mathbf{X}] &= \delta_a^{2001} \\
\mathbb{E}\left[b_1(A, U_1) - b_0(A, U_0)\right|\mathbf{X}] &= \delta_b^{2001} \\
\mathbb{E}\left[c_1(A, U_1) - c_0(A, U_0)\right|\mathbf{X}] &= \delta_c^{2001}
\end{aligned}$$

with an analogous restriction holding across periods 1 and 2. This specification allows for the demand elasticity to shift over time (albeit in a way that is homogenous – in a functional sense – across households). The 2000 to 2002 period coincided with the 'coffee crisis' in Nicaragua, so there is some *a priori* reason to believe that macro-shifts in the demand elasticity may be important.[17]

Defining

$$\mathbf{W} = \begin{pmatrix} 0 & 0 \\ X_1' & 0 \\ 0 & X_2' \end{pmatrix},$$

gives, under marginal stationarity, a general semiparametric regression model of

$$\mathbb{E}[\mathbf{Y}|\mathbf{W},\mathbf{X}] = \mathbf{W}\boldsymbol{\delta} + \mathbf{X}\boldsymbol{\beta}(\mathbf{X}),$$

with $\boldsymbol{\delta} = \left(\delta_a^{2001}, \delta_b^{2001}, \delta_c^{2001}, \delta_a^{2002}, \delta_b^{2002}, \delta_c^{2002}\right)'$ and

$$\begin{aligned} \boldsymbol{\beta}(\mathbf{x}) &= (\mathbb{E}[a_0\left(A,U_t\right)|\, X=x], \mathbb{E}[b_0\left(A,U_t\right)|\, X=x], \mathbb{E}[c_0\left(A,U_t\right)|\, X=x])' \\ &= \left(\alpha\left(x\right), \beta\left(x\right), \gamma\left(x\right)\right)'. \end{aligned}$$

Relative to prior work, the distinguishing feature of the above model is that it allows for the elasticity of calorie demand to vary across households and time in a way that may co-vary with total outlay. The specification also allows a household's elasticity to *structurally* vary with its income, albeit in a restricted way. Subramanian and Deaton (1996), in contrast, allow a household's calorie elasticity to structurally vary with income in a fully nonparametric way. However they assume that this nonlinear mapping is homogenous across all households.

Below we compare estimates of $\bar{\epsilon}_t(\mathrm{q}) = \mathbb{E}\left[\epsilon_t(\mathrm{q};\, A, U_t)\right]$ with those derived from conventional panel data estimators as well as the (now standard) cross-sectional local linear regression estimator popularized in this context by Subramanian and Deaton (1996). As noted above, the elasticity estimate based upon the conventional within-household regression estimator will – loosely speaking – overemphasize those households with large year-to-year changes in total expenditure. In poor village economies, like those from which our data are drawn, the ability to smooth consumption over time can vary significantly across households and therefore the within-household elasticity estimate may be far from the desired population average.

Nonparametric estimates based on cross-section data may also be affected by correlated household heterogeneity. Subramanian and Deaton (1996), in the absence of panel data, use a single cross-section to estimate the calorie elasticity.[18] Their estimate is the sample analog of the deriv-

[17]Skoufias (2003) finds that the estimated calorie demand elasticity is largely insensitive to aggregate price changes in Indonesia.

[18]We focus on Subramanian and Deaton (1996) paper as its basic modelling approach has become prototypical in this literature (e.g., Skoufias, 2003, Logon, 2006, Smith and Subandoro, 2007). Behrman and Deolalikar (1987), who work with a small two-period panel, use conventional linear fixed-effects methods to study the demand for calories.

ative of the conditional expectation function $\mathbb{E}\left[\ln(\mathrm{Cal}_0)\right|\ln(\mathrm{Exp}_0)=\mathrm{q}]$, which under (20) equals

$$\frac{\partial\mathbb{E}\left[\ln(\mathrm{Cal}_0)\right|\ln(\mathrm{Exp}_0)=\mathrm{q}]}{\partial\mathrm{q}}=\beta\left(\mathrm{q}\right)-\frac{\gamma\left(\mathrm{q}\right)}{e^{\mathrm{q}}}+\left\{\nabla_q\beta\left(\mathrm{q}\right)\mathrm{q}+\frac{\nabla_q\gamma\left(\mathrm{q}\right)}{e^{\mathrm{q}}}\right\},$$

where, in an abuse of our established notation, $\beta\left(\mathrm{q}\right)=\mathbb{E}\left[b_0\left(A,U_t\right)\right|\ln(\mathrm{Exp}_0)=\mathrm{q}]$ and $\gamma\left(\mathrm{q}\right)=\mathbb{E}\left[c_0\left(A,U_t\right)\right|\ln(\mathrm{Exp}_0)=\mathrm{q}]$. The first term is structural, while the second reflects heterogeneity bias. If $\beta\left(\mathrm{q}\right)$ and $\gamma\left(\mathrm{q}\right)$ vary with income in the population, then the cross-sectional kernel regression estimator will be inconsistent.

Our point is not to argue for the superiority of one approach or the other, but to highlight the substantive differences between then. At the cost of a parametric form for the household's elasticity, our method allows for substantial correlated heterogeneity across households. Subramanian and Deaton (1996), while allowing for nonlinearity in the structural mapping from (log) income to (log) calories, assumes substantial homogeneity across households. Richer panel data could lessen these trade-offs.

5.2 Data description and overview

We use data collected in conjunction with an external evaluation of the Nicaraguan conditional cash transfer program Red de Protección Social (RPS) (see IFPRI, 2005). The RPS evaluation sample is a panel of 1,581 households from 42 rural communities in the departments of Madriz and Matagalpa, located in the northern part of the Central Region of Nicaragua. Twenty one of the sampled communities were randomly assigned to participate in the RPS program. Each sampled household was first interviewed in August/September 2000 with follow-ups attempted in October of both 2001 and 2002. Here we analyze a balanced panel of 1,358 households from all three waves.[19]

The survey was fielded using an abbreviated version of the 1998 Nicaraguan Living Standards Measurement Survey (LSMS) instrument. As such it includes a detailed consumption module with information on household expenditure, both actual and in kind, on 59 specific foods and several dozen other common budget categories (e.g., housing and utilities, health, education, and household goods). The responses to these questions were combined to form an annualized consumption aggregate, C_{it}. In forming this variable we followed the algorithm outlined by Deaton and Zaidi (2002).

In addition to recording food expenditures, actual quantities of foods acquired are available. Using conversion factors listed in the World Bank (2002) and Instituto Nacional de Estadísticas y Censos (2005) (henceforth INEC) we converted all food quantities into grams. We then used the caloric content and edible percent information in the Instituto de Nutrición de Centro América y Panamá (2000) (henceforth INCAP) food composition tables to construct a measure of daily total

[19] A total of 1,359 households were successfully interviewed in all three waves. One of these households reports zero food expenditures (and hence caloric availability) in one wave and is dropped from our sample. The preparation of our estimation sample from the raw public release data files involved some complex and laborious data-processing. We outline the procedures used in this section. A sequence of heavily commented STATA do files, which read in the IFPRI (2005) release of the data and output a text file of our estimation sample will eventually be made available online at `http://www.econ.berkeley.edu/~bgraham/`.

calories available for each household.[20] The logarithm of this measure divided by household size, Y_{it}, serves as the dependent variable in our analysis.

The combination of both expenditure and quantity information at the household-level also allowed us to estimate unit prices for foods. These unit values were used to form a Paasche cost-of-living index for the i^{th} household in year t of

$$I_{it} = \left[S_{it} \left\{ \sum_{f=1}^{F} W_{f,it} \left(P_f^b / P_{f,it} \right) \right\} + (1 - S_{it}) J_{it} \right]^{-1}, \tag{22}$$

where S_{it} is the fraction of household spending devoted to food, $W_{f,it}$ the share of overall food spending devoted to the f^{th} specific food, $P_{f,it}$ the year t unit price paid by the household for food f, and P_f^b its 'base' price (equal to the relevant 2001 sample median price). We use 2001 as our base year since it facilitates comparison with information from a nationwide LSMS survey fielded that year. Following the suggestion of Deaton and Zaidi (2002) we replace household-level unit prices with village medians in order to reduce noise in the price data. In the absence of price information on nonfood goods we set J_{it} equal to one in 2001 and to the national consumer price index (CPI) in 2000 and 2002. Our independent variable of interest is real per capita consumption in thousands of Cordobas: $\text{Exp}_{it} = ([C_{it}/I_{it}]/1,000)/M_{it}$; M_{it} is total household size.

Tables 1, 2 and 3 summarize some key features of our estimation sample. Panel A of Tables 1 give the share of total food spending devoted to each of eleven broad food categories. Spending on staples (cereals, roots and pulses) accounts for about half of the average household's food budget and over two thirds of its calories (Tables 1 and 2). Among the poorest quartile of households an average of around 55 percent of budgets are devoted to, and over three quarters of calories available derived from, staples. Spending on vegetables, fruit and meat accounts for less than 15 percent of the average household's food budget and less than 3 percent of calories available. That such a large fraction of calories are derived from staples, while not good dietary practice, is not uncommon in poor households elsewhere in the developing world (cf., Subramanian and Deaton, 1996; Smith and Subandoro, 2007).

Panel B of the table lists real annual expenditure in Cordobas per adult equivalent and per capita. Adult equivalents are defined in terms of age- and gender-specific FAO (2001) recommended energy intakes for individuals engaging in 'light activity' relative to prime-aged males. As a point of reference the 2001 average annual expenditure per capita across all of Nicaragua was a nominal C\$7,781, while amongst rural households it was C\$5,038 (World Bank. 2003). The 42 communities in our sample, consistent with their participation in an anti-poverty demonstration experiment, are considerably poorer than the average Nicaraguan rural community.[21]

[20] In forming our measure of calorie availability we followed the general recommendations of Smith and Subandoro (2007).

[21] In October of 2001 the Coroba-to-US\$ exchange rate was 13.65. Therefore per capita consumption levels in our sample averaged less than US\$ 300 per year.

Panel A:	Expenditure Shares (%)								
	All			Lower 25%			Upper 25%		
	2000	2001	2002	2000	2001	2002	2000	2001	2002
Cereals	49.1	36.0	32.7	53.3	40.9	35.7	45.7	31.6	29.4
Roots	1.3	3.1	2.7	1.3	2.6	2.0	1.5	3.6	3.6
Pulses	11.6	12.5	13.6	11.2	13.8	16.5	10.6	10.7	11.3
Vegetables	3.2	4.9	4.5	2.8	4.3	3.4	3.8	5.8	5.3
Fruit	0.6	0.9	1.1	0.5	0.7	0.9	0.8	1.2	1.2
Meat	3.1	6.9	7.7	2.2	4.0	5.1	5.3	9.9	10.4
Dairy	11.2	14.7	17.3	9.0	12.0	15.0	13.1	16.8	19.2
Oil	4.0	5.0	5.0	3.5	5.2	5.0	3.9	4.7	4.7
Other foods	15.8	16.0	15.4	16.2	16.7	16.5	15.4	15.7	14.9
Staples$^{\diamond}$	62.1	51.6	49.0	65.8	57.3	54.1	57.8	45.9	44.3
Panel B:	Total Real Expenditure & Calories								
Expenditure per adult$^{\flat}$	5,506	4,679	4,510	2,503	2,397	2,200	9,481	7,578	7,460
(Expenditure per capita)	(4,277)	(3,764)	(3,887)	(2,016)	(2,130)	(2,102)	(7,302)	(5,845)	(6,114)
Food share	71.2	69.2	68.8	73.8	69.1	68.6	67.0	67.9	67.6
Calories per adult$^{\flat}$	2,701	3,015	2,948	1,706	2,127	2,013	3,738	3,849	3,758
(Calories per capita)	(2,086)	(2,435)	(2,529)	(1,350)	(1,854)	(1,873)	(2,842)	(2,962)	(3,041)
Percent energy deficient$^{\natural}$	51.0	39.3	39.7	85.0	69.7	76.2	19.8	14.5	13.0

Table 1: Real food expenditure budget shares of RPS households from 2000 to 2002

NOTES: Authors' calculations based on a balanced panel of 1,358 households from the RPS evaluation dataset (see IFPRI (2005)). Real household expenditure equals total annualized nominal outlay divided by a Paasche cost-of-living index. Base prices for the price index are 2001 sample medians. The nomial exchange rate in October of 2001 was 13.65 Cordobas per US dollar. Total calorie availability is calculated using the RPS food quantity data and the calorie content and edible portion information contained in INCAP (2000). Lower and upper 25 percent refers to the bottom and top quartiles of households based on the average of year 2000, 2001 and 2002 real consumption per adult equivalent and thus contains the same set of households in all three years.

$^{\diamond}$ Sum of cereal, roots and pulses.

$^{\flat}$ "Adults" correspond to adult equivalents based on FAO (2001) recommended energy requirements for light activity.

$^{\natural}$ Percentage of households with estimated calorie availability less than FAO (2001) recommendations for light activity given household demographics.

	Calorie Shares (%)								
	All			Lower 25%			Upper 25%		
	2000	2001	2002	2000	2001	2002	2000	2001	2002
Cereals	57.7	60.3	59.9	60.1	63.9	62.0	55.5	57.1	57.4
Roots	1.5	1.5	1.6	1.9	1.5	1.2	1.6	1.8	2.1
Pulses	13.1	11.3	12.8	12.1	11.3	13.3	13.1	11.0	12.1
Vegetables	0.7	0.7	0.6	0.6	0.6	0.4	0.8	0.9	0.8
Fruit	0.3	0.5	0.4	0.3	0.3	0.4	0.5	0.7	5.8
Meat	0.7	1.3	1.3	0.5	0.7	0.7	1.3	1.9	1.9
Dairy	4.1	4.3	4.5	3.4	3.0	3.4	4.7	5.2	5.5
Oil	6.9	7.6	7.5	5.8	6.9	6.7	7.4	8.1	8.0
Other foods	15.0	12.6	11.4	14.7	11.9	11.9	15.2	13.2	11.5
Staples$^{\diamond}$	72.3	73.1	74.3	74.7	76.7	76.6	70.2	69.9	71.7

Table 2: Calorie shares of RPS households from 2000 to 2002
NOTES: Authors' calculations based on a balanced panel of 1,358 households from the RPS evaluation dataset (see IFPRI (2005)). Total calorie availability is calculated using the RPS food quantity data and the calorie content and edible portion information contained in INCAP (2000). Lower and upper 25 percent refers to the bottom and top quartiles of households based on the average of year 2000, 2001 and 2002 real consumption per adult equivalent and thus contains the same set of households in all three years.
$^{\diamond}$ Sum of cereal, roots and pulses.

Using the FAO (2001) energy intake recommendations for 'light activity' we categorized each household, on the basis of its demographic structure, as energy deficient, or not. By this criterion approximately 40 percent of households in our sample are energy deficient each period. Amongst the poorest quartile this fraction rises to over 75 percent. These figures are reported in Panel B of Table 1.

Table 3 reports the median amount of Cordobas paid per one thousand calories by food type and expenditure quartiles. As found in other parts of the developing world, 'rich' households spend more per calorie than poor households, however, these price differences are not especially large in our sample. If quality upgrading is an important feature of food demand, then the elasticity of calorie demand with respect to total expenditure may be quite low even if the elasticity of food expenditure is quite high (Behrman and Deolalikar, 1987; Subramanian and Deaton, 1996).

5.3 Computation

For computation we employ an 'instrumental variables' procedure. Our estimates of π, $\boldsymbol{\delta}$, $\boldsymbol{\beta}^S$ and $\boldsymbol{\beta}^M$ are given by the solution to (suppressing the dependence of our estimator on the choice of bandwidth)

$$\sum_{i=1}^{N} \psi_i(\widehat{\theta})/N = 0,$$

	Median Cordobas Paid per 1,000 calories								
	All			Bottom 25%			Top 25%		
	2000	2001	2002	2000	2001	2002	2000	2001	2002
Cereals	2.3	1.3	1.2	2.1	1.2	1.0	2.7	1.5	1.4
Roots	5.7	8.6	7.2	3.0	7.2	6.0	6.5	8.6	7.2
Pulses	2.6	2.6	2.3	2.6	2.6	2.4	2.6	2.6	2.6
Vegetables	20.5	23.2	22.7	17.2	22.5	19.6	22.5	24.2	23.5
Fruit	6.6	6.3	6.6	5.4	5.1	5.3	6.6	6.6	6.7
Meat	18.2	19.1	18.6	15.5	18.6	18.6	18.5	20.1	19.3
Dairy	9.3	10.1	10.0	9.1	10.6	10.4	9.7	10.2	10.1
Oil	1.6	1.5	1.5	1.6	1.6	1.5	1.6	1.5	1.5
Other foods	3.0	3.1	3.1	3.0	2.9	2.7	3.2	3.5	3.6
All foods	3.0	2.4	2.4	2.6	2.0	2.0	3.5	3.0	2.9

Table 3: Real Cordobas spent by RPS households from 2000 to 2002 per 1,000 calories available by food category
NOTES: Authors' calculations based on a balanced panel of 1,358 households from the RPS evaluation dataset (IFPRI, 2005). Reported calorie prices are the median among households with positive consumption in the relevant category. Lower and upper 25 percent refers to the bottom and top quartiles of households based on the average of year 2000, 2001 and 2002 real consumption per adult equivalent. See notes to Table 1 for additional details.

with $\widehat{\theta} = \left(\widehat{\pi}, \widehat{\boldsymbol{\delta}}', \widehat{\boldsymbol{\beta}}^{S\prime}, \widehat{\boldsymbol{\beta}}^{M\prime}\right)'$ and

$$\psi_i(\theta) = \begin{pmatrix} \mathbf{1}(\left|\tilde{X}_i\right| \leq h_N) - \pi \\ \mathbf{Z}_i' \left(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{W}}_i \boldsymbol{\delta} - \tilde{X}_i \left[\begin{pmatrix} \mathbf{1}(\left|\tilde{X}_i\right| \leq h_N) & \mathbf{1}(\left|\tilde{X}_i\right| > h_N) \end{pmatrix} \otimes \mathbf{I}_p\right] \begin{pmatrix} \boldsymbol{\beta}^S \\ \boldsymbol{\beta}^M \end{pmatrix}\right) \end{pmatrix},$$

where

$$\tilde{\mathbf{Y}}_i = \mathbf{X}_i^* \mathbf{Y}_i, \qquad \tilde{\mathbf{W}}_i = \mathbf{X}_i^* \mathbf{W}_i,$$

and the $T \times q \times 2p$ instrument matrix is given by

$$\mathbf{Z}_i \equiv \left[\mathbf{1}(\left|\tilde{X}_i\right| \leq h_N) \cdot \tilde{\mathbf{W}}, \mathbf{1}(\left|\tilde{X}_i\right| \leq h_N) \cdot \tilde{X}_i \cdot \mathbf{I}_p, \frac{\mathbf{1}(\left|\tilde{X}_i\right| > h_N)}{\tilde{X}_i} \cdot \mathbf{I}_p\right].$$

This procedure is numerically equivalent to the two-step procedure described above where in the first step $\widehat{\boldsymbol{\delta}}$ and $\widehat{\boldsymbol{\beta}}^S$ are the local linear estimates

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\left|\tilde{X}_i\right| \leq h_N) \begin{pmatrix} \tilde{\mathbf{W}}' \\ \tilde{X}_i \cdot \mathbf{I}_p \end{pmatrix} \left(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{W}}_i \widehat{\boldsymbol{\delta}} - \tilde{X}_i \widehat{\boldsymbol{\beta}}^S\right) = 0,$$

with $\widehat{\boldsymbol{\beta}}^M$ then given by

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\left|\tilde{X}_i\right| > h_N)\left\{\mathbf{X}_i^{-1}\left(\mathbf{Y}_i - \mathbf{W}_i\widehat{\boldsymbol{\delta}}\right) - \widehat{\boldsymbol{\beta}}^M\right\} = 0.$$

The distribution of X_t does not appear to have discrete components, therefore $\widehat{\boldsymbol{\beta}}^M$ should provide – in principal – the basis for a consistent estimate of the average elasticity of calorie demand. However, due to the extreme within-unit colinearity of $\ln(\text{Exp}_t)$ and Exp_t^{-1} the density of $\det(\mathbf{X}_i)$ is substantial in the neighborhood of zero (cf., Figure 2). The extreme 'irregularity' of our application, with many 'near stayers' necessitates substantial trimming. For this we reason report, and prefer, the estimate

$$\widehat{\boldsymbol{\beta}} = \widehat{\pi}\widehat{\boldsymbol{\beta}}^S + (1-\widehat{\pi})\,\widehat{\boldsymbol{\beta}}^M.$$

The year 2000 estimated average elasticity of calorie demand is then given by

$$\widehat{\epsilon}_0(\mathrm{q}) = \widehat{\beta} - \widehat{\gamma}\mathrm{q}^{-1},$$

with the year 2001, 2002 elasticities also incorporating the relevant elements of $\widehat{\boldsymbol{\delta}}$. We calculate average elasticities for total outlay equal to the 25th, 50th and 75th percentiles of the 2000 expenditure distribution (respectively q equal to 2.267, 3.650 and 5.539 thousands of Cordobas).

We compute standard errors using the conventional GMM variance-covariance estimator and the delta method (cf., Newey and McFadden, 1994). In doing so we only assume independence across villages, not between them. While this estimator includes estimates of terms that are asymptotically negligible, it is computationally convenient and may lead to confidence intervals with better coverage properties in small samples.

To select h_N we employ a variant of 'k-fold' cross-validation, choosing h_N to minimize

$$CV(h_N) \equiv \sum_{i=1}^{N}\psi_i(\widehat{\theta}_{-i})'\psi_i(\widehat{\theta}_{-i})/N,$$

where $\widehat{\theta}_{-i}$ is the estimate calculated by omitting the i^{th} observation *and all other households in their village*. We explore the sensitivity of our estimates to this choice of bandwidth.

5.4 Results

Table 4 reports pooled OLS and linear fixed effects estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$. Column (a) of each panel estimates models without aggregate time effects. Columns (b) and (c) of each panel respectively allow for the intercept, and the intercept and slope coefficients, to shift across periods. These aggregate time effects are jointly significant in all specifications. The implied elasticities for each year, evaluated at each of three total expenditure levels, are reported in Table 5.

The pooled OLS and FE elasticities are similar in magnitude providing little evidence of significant heterogeneity bias. There is evidence that the elasticity of calorie demand declines with total

	Pooled OLS			FE-OLS		
	(1.a)	(1.b)	(1.c)	(2.a)	(2.b)	(2.c)
β	0.4450	0.4812	0.4741	0.4629	0.5020	0.4875
	(0.0414)	(0.0384)	(0.0312)	(0.0423)	(0.0393)	(0.0374)
γ	-0.4238	-0.3640	-0.5514	-0.4010	-0.3391	-0.5052
	(0.1450)	(0.1325)	(0.0780)	(0.1376)	(0.1246)	(0.0835)
	Aggregate Time Effects			Aggregate Time Effects		
δ_a^{2001}	–	0.2385	0.1418	–	0.2400	0.1907
		(0.0400)	(0.1706)		(0.0401)	(0.1501)
δ_a^{2002}	–	0.2757	0.0188	–	0.2769	-0.0206
		(0.0284)	(0.1324)		(0.0286)	(0.1378)
δ_b^{2001}	–	–	0.0033	–	–	-0.0111
			(0.0734)			(0.0683)
δ_b^{2002}	–	–	0.0700	–	–	0.1002
			(0.0635)			(0.0681)
δ_c^{2001}	–	–	0.2575	–	–	0.1716
			(0.2381)			(0.1998)
δ_c^{2002}	–	–	0.4770	–	–	0.4873
			(0.1654)			(0.1604)
p-value $H_0 : \delta = 0$	–	0.000	0.000	–	0.000	0.000

Table 4: Conventional parametric estimates of the elasticity of calorie demand with respect to household expenditure

NOTES: Estimates based on the balanced panel of 1,358 households described in the main text. "Pooled OLS" denotes least squares applied to the pooled 2000, 2001 and 2002 samples, "FE-OLS" denotes least squares estimates with household-specific intercepts. The standard errors are computed in a way that allows for arbitrary within-village correlation in disturbances across households and time.

	Pooled OLS			FE-OLS			Kernel
	(1.a)	(1.b)	(1.c)	(2.a)	(2.b)	(2.c)	3
			At 25th percentile				
2000	0.6320 (0.0345)	0.6408 (0.0317)	0.7173 (0.0289)	0.6398 (0.0284)	0.6516 (0.0242)	0.7104 (0.0279)	0.6683 (0.5770, 0.7640)
2001	–	–	0.6070 (0.0582)	–	–	0.6235 (0.0475)	0.6842 (0.5831, 0.8012)
2002	–	–	0.5769 (0.0359)	–	–	0.5956 (0.0279)	0.6640 (0.5653, 0.7641)
			At 50th percentile				
2000	0.5612 (0.0207)	0.5799 (0.0195)	0.6252 (0.0247)	0.5728 (0.0177)	0.5949 (0.0154)	0.6259 (0.0264)	0.6548 (0.5812, 0.7213)
2001	–	–	0.5579 (0.0360)	–	–	0.5679 (0.0312)	0.5938 (0.5190, 0.6687)
2002	–	–	0.5304 (0.0371)	–	–	0.5926 (0.0266)	0.5110 (0.4507, 0.5736)
			At 75th percentile				
2000	0.5216 (0.0226)	0.5459 (0.0214)	0.5736 (0.0251)	0.5353 (0.0223)	0.5632 (0.0205)	0.5787 (0.0286)	0.6377 (0.5704, 0.6968)
2001	–	–	0.5645 (0.0280)	–	–	0.5366 (0.0336)	0.4971 (0.4326, 0.5701)
2002	–	–	0.5575 (0.0346)	–	–	0.5909 (0.0362)	0.4378 (0.3709, 0.5074)

Table 5: Parametric and nonparametric calorie demand elasticities
NOTES: Elasticities reported in Columns 1.a-1.c and 2.a to 2.c calculated using the estimates reported in Table 4. The reported standard errors are computed using the delta method. The column 3 'Kernel' elasticity estimates are based on three cross-sectional local linear kernel regressions of the type employed by Subramanian and Deaton (1996). As in that work the bandwith was selected informally by the 'eye'. Below the reported elasticities are 95 percent confidence intervals based on 1,000 bootstrap replications (with the village being treated as the relevant sampling unit).

outlay; γ is significantly negative in all specifications. How the magnitude of the decline is modest across the interquartile range of the 2000 household expenditure distribution. These estimates are on the higher end of those reported in the literature, but not implausible given that our sample was selected for its extreme poverty (cf., Strauss and Thomas, 1995, Table 34.1).

Table 5 also reports elasticity estimates based on three, corresponding to the years 2000, 2001 and 2002, cross-sectional local linear regressions of $\ln(\mathrm{Cal}_t)$ onto $\ln(\mathrm{Exp}_t)$. This approach to estimating the calorie demand elasticity was used by Subramanian and Deaton (1996) and is now virtually standard in the literature. The local linear elasticity estimates differ little from their pooled OLS or FE-OLS counterparts. Figure 1 plots the coefficient on $\ln(\mathrm{Exp}_t) - \mathrm{q}$ for a grid of equally-spaced values of q between the 5th and 95th percentiles of the 2000 total outlay distribution. While the figure shows clear evidence of a demand elasticity that declines with total outlay, this decline is relatively muted across the interquartile range of the 2000 total outlay distribution. In summary the pooled OLS, fixed effects and cross-sectional nonparametric regression estimates of the calorie

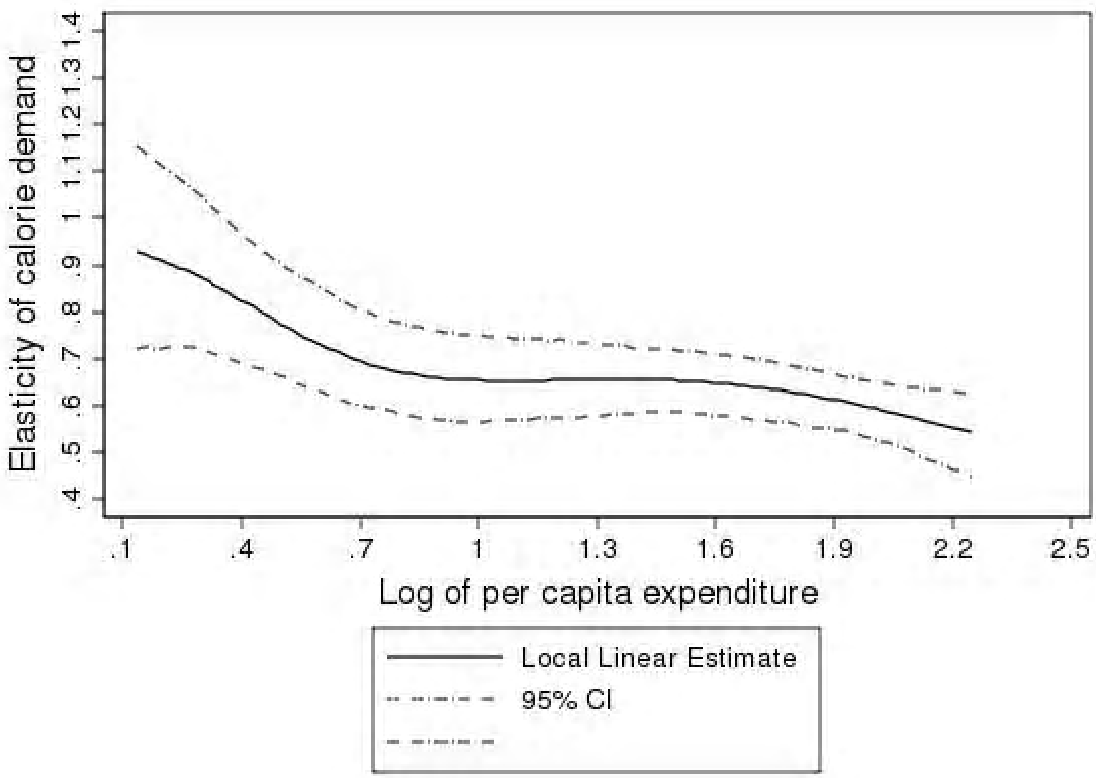

Figure 1: Local linear regression estimates of the calorie elasticity using the year 2000 cross-section. NOTES: Estimates based on a Gaussian kernel with a bandwidth selected by the 'eye'. Confidence bands based on 1,000 bootstrap replications with the village being treated as the sampling unit. The elasticity is estimated at a grid of equally-spaced values between the 5th and 95th percentiles of the 2000 total outlay distribution.

demand elasticity are all rather similar.

Table 6 reports estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ based on our CRC model. The corresponding elasticity estimates are reported in Table 7. Column 1 is the simple 'Mundlak/Chamberlain' estimator $\widehat{\beta}^M = N^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\mathbf{X}_i'\mathbf{Y}_i\right)$. Consistency of this estimator requires that $\mathbb{E}[1/\det\left(\mathbf{X}_i'\mathbf{X}_i\right)] < \infty$. Given that the density of $\det\left(\mathbf{X}\right)$ is substantial in the neighborhood of zero (see Figure 2), this condition is unlikely to be satisfied in the present setting. The degree of 'irregularity' in our application is confirmed by the nonsensical elasticity estimates produced by this estimator.

Columns 2 through 4 report estimates using our procedure based on models with, respectively, no time effects, intercept shifts only, and intercept and slope shifts together. In each column the bandwidth is selected by the cross-validation procedure described above (as an example the cross-validation criterion for the Column 3 specification is plotted in Figure 3). We focus on the Column 4 estimates in the discussion which follows.

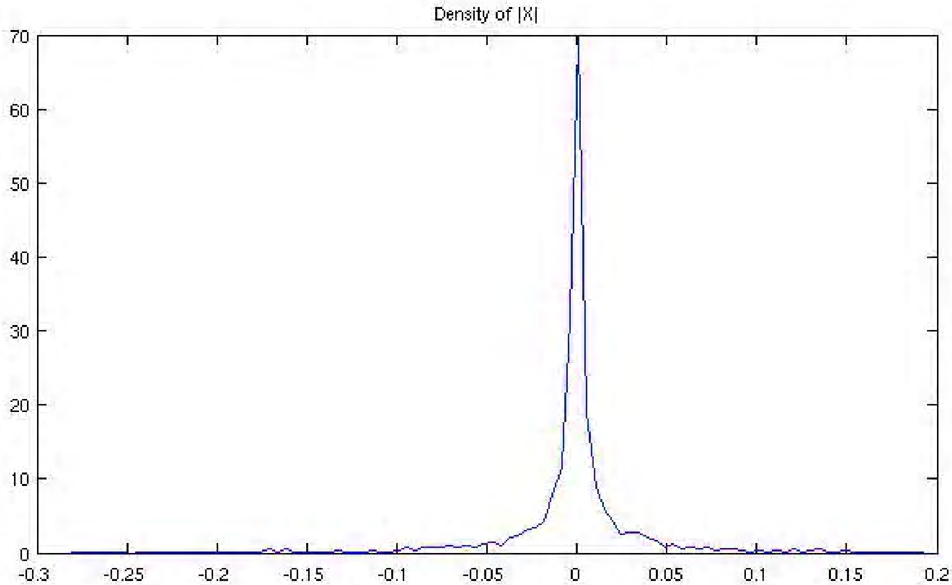

Figure 2: Kernel density estimate of the density of $\det(\mathbf{X})$.
NOTES: Density estimate based on a Gaussian kernel and Silverman's rule-of-thumb bandwidth. The top and bottom 1 percentiles of the empirical $\det(\mathbf{X})$ distribution were excluded from the estimates in order to produce a more viewable figure.

Comparing the Column 4 elasticity estimates reported in Table 7 with their counterparts in Table 5 (Column 3, 6, and 7), we see that the CRC elasticities decline more sharply with total outlay, being higher at the 25 percentile of the outlay distribution and lower at the 75 percentile across all three years (although sampling error suggests some caution in pushing this result too far). Furthermore our elasticity estimates are reasonably insensitive to modest variation in the bandwidth. Columns 5 through 8 reported estimates based on a bandwidth equal to 1/8, 1/4, 1/2 and twice the Column 4 value. For smaller bandwidths the decline of the average elasticity with total outlay is somewhat more pronounced, while for large bandwidths it is less so.

5.5 Summary and extensions

In this paper we have outlined a new estimator for the correlated random coefficients panel data model of Chamberlain (1992a). Our estimator is designed for situations where the regularity conditions required for his method-of-moments procedure fail. We illustrate the use of our methods in an exploration of the elasticity of demand for calories in a sample of poor Nicaraguan households. This application is highly 'irregular', with many 'near stayers' in the sample. This implies that (i) the elasticity estimates based on the textbook linear FE estimator may far from the relevant population average and (ii) the use of Chamberlain's (1992a) estimator is inappropriate. Both these facts motivate the use of our trimmed estimator. Our estimates are suggestive, albeit not decisively so, of the presence of correlated random coefficient heterogeneity.

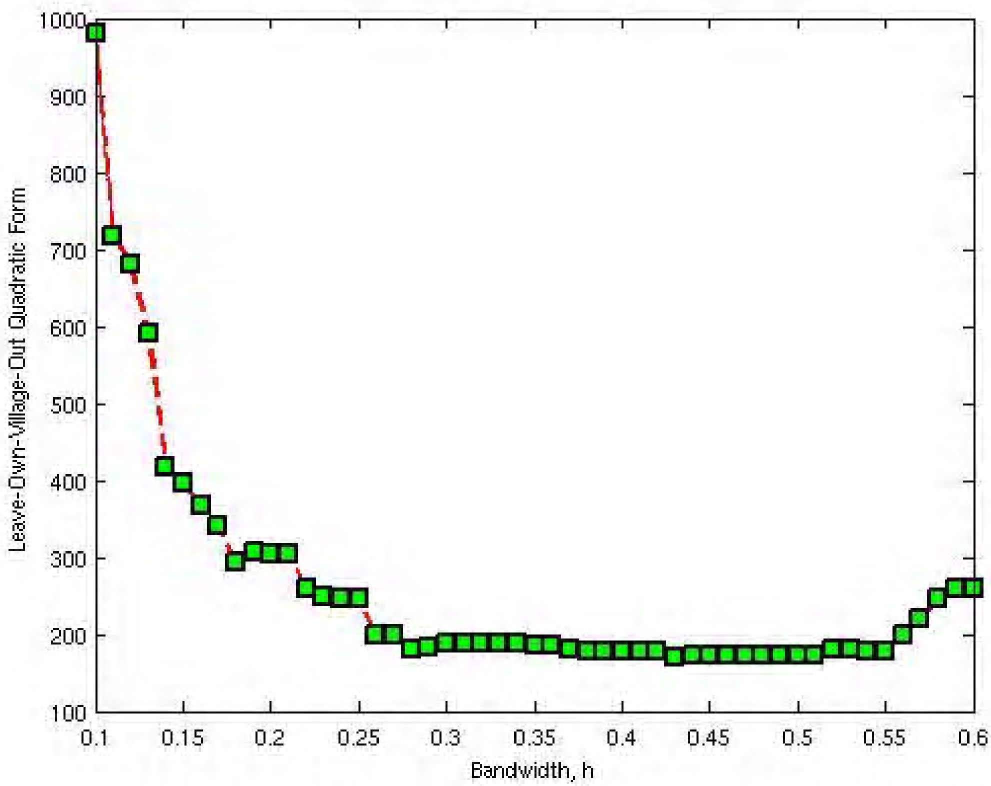

Figure 3: Leave-own-village-out 'IV' quadratic form for different bandwidth values – Column 3, Table 6 specification.

While our procedure is simple to implement, it does require choosing a smoothing parameter. As in other areas of the semiparametric estimation literature, our theory provides little guidance on this choice. In ongoing work we are studying how to extend our methods to estimate quantile partial effects (e.g., unconditional quantiles of the distribution of $\mathbf{b}(A, U_t)$) and to accommodate additional 'triangular endogeneity'.

	CRC-Panel							
	(1)	(2)	(3)	(4)	(5)	(5)	(6)	(8)
β^S	–	0.2551 (0.2476)	0.3319 (0.2281)	0.4395 (0.2139)	0.0487 (0.3779)	0.1963 (0.3342)	0.2633 (0.2942)	0.4299 (0.2051)
γ^S	–	−1.0691 (0.4675)	−0.9091 (0.4525)	−0.5474 (0.4331)	−1.7739 (0.9018)	−1.5061 (0.7657)	−1.0588 (0.6849)	0.0901 (0.3628)
β^M	3.0498 (13.9910)	0.3914 (0.4154)	0.8664 (0.4070)	1.0193 (0.3940)	0.3007 (0.2667)	0.4421 (0.2428)	0.8696 (0.3940)	0.3727 (0.7848)
γ^M	29.4492 (57.3426)	−0.5191 (0.6876)	0.0477 (0.5900)	0.7384 (0.5809)	−0.9828 (0.5329)	−0.5454 (0.5615)	0.0993 (0.4312)	−0.1069 (0.7933)
β	–	0.2565 (0.2430)	0.3354 (0.2434)	0.4447 (0.2936)	0.0765 (0.3759)	0.2090 (0.3880)	0.2771 (0.4505)	0.7491 (0.2105)
γ	–	−1.0634 (0.4940)	−0.9027 (0.4939)	−0.5361 (0.6082)	−1.6865 (1.0328)	−1.4566 (1.1023)	−1.0323 (0.9768)	0.0897 (0.3644)
	Aggregate Time Effects							
δ_a^{2001}	–	–	0.2794 (0.0481)	1.2888 (0.4056)	0.6010 (0.3408)	0.6276 (0.3562)	0.9125 (0.3857)	1.2709 (0.3922)
δ_a^{2002}	–	–	0.2724 (0.0422)	0.9399 (0.3660)	0.3649 (0.3249)	0.5258 (0.3477)	0.7699 (0.3431)	0.8616 (0.3386)
δ_b^{2001}	–	–	–	−0.4816 (0.2002)	−0.1752 (0.1512)	−0.1931 (0.1634)	−0.3383 (0.1835)	−0.4697 (0.1943)
δ_b^{2002}	–	–	–	−0.3327 (0.1706)	−0.0745 (0.1501)	−0.1433 (0.1605)	−0.2575 (0.1623)	−0.3041 (0.1625)
δ_c^{2001}	–	–	–	−1.1981 (0.4304)	−0.3995 (0.4490)	−0.3755 (0.4379)	−0.6614 (0.4590)	−1.2180 (0.4216)
δ_c^{2002}	–	–	–	−0.7171 (0.4337)	0.0299 (0.4269)	−0.1435 (0.4290)	−0.5196 (0.4030)	−0.6269 (0.3942)
p-value $H_0 : \delta = 0$	–	–	0.000	0.000	0.000	0.000	0.000	0.000
h_N	0	0.380	0.520	0.430	0.054	0.108	0.215	0.860
Percent 'trimmed'	0	99.0	99.3	99.1	89.0	94.8	97.7	99.8

Table 6: Correlated random coefficient estimates of the elasticity of calorie demand with respect to household expenditure
NOTES: Estimates based on the balanced panel of 1,358 households described in the main text. The Column 1 estimates correspond to a 'naive' application of Chamberlain's (1992) estimator with no trimming. Leave-Own-Village-Out cross validation (as described in the main text) is used to select h_N in Columns 2 to 4. Columns 5 to 8 set h_N equal to 1/8, 1/4, 1/2 and 2 times its column 4 value. The standard errors are computed in a way that allows for arbitrary within-village correlation in disturbances across households and time.

	CRC Estimates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				At 25th percentile				
2000	−9.9423	0.7256	0.7337	0.6812	0.8205	0.8516	0.7326	0.7096
	(15.6502)	(0.1069)	(0.0533)	(0.0749)	(0.1751)	(0.1841)	(0.0973)	(0.1112)
2001	–	–	–	0.7281	0.8216	0.8242	0.6860	0.7772
				(0.0752)	(0.1913)	(0.1846)	(0.1167)	(0.1034)
2002	–	–	–	0.6649	0.7328	0.7716	0.7043	0.6820
				(0.0943)	(0.1639)	(0.1806)	(0.1104)	(0.0977)
				At 50th percentile				
2000	−5.0196	0.5479	0.5828	0.5916	0.5386	0.6081	0.5600	0.7245
	(8.6997)	(0.1353)	(0.1142)	(0.1385)	(0.1541)	(0.1498)	(0.1977)	(0.1356)
2001	–	–	–	0.4382	0.4729	0.5179	0.4029	0.5886
				(0.1350)	(0.1607)	(0.1556)	(0.2176)	(0.1583)
2002	–	–	–	0.4554	0.4559	0.5041	0.4448	0.5922
				(0.1039)	(0.1605)	(0.1422)	(0.1951)	(0.1427)
				At 75th percentile				
2000	−2.2671	0.4485	0.4984	0.5414	0.3810	0.4719	0.4635	0.7329
	(7.8200)	(0.1676)	(0.1572)	(0.1892)	(0.2141)	(0.2136)	(0.2808)	(0.1581)
2001	–	–	–	0.2761	0.2779	0.3467	0.2446	0.4831
				(0.1899)	(0.2145)	(0.2105)	(0.2934)	(0.2000)
2002	–	–	–	0.3382	0.3011	0.3546	0.2998	0.5420
				(0.1419)	(0.2251)	(0.2007)	(0.2662)	(0.1779)

Table 7: Correlated random coefficient calorie demand elasticities
NOTES: Elasticities calculated using estimates reported in Table 6. The standard errors are computed using the delta method.

References

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

Altonji, Joseph G. and Thomas A. Dunn. (1996). "Using siblings to estimate the effect of school quality on wages," *Review of Economics and Statistics* 78 (4): 665 - 671.

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.

Arellano, Manuel and Stéphane Bonhomme. (2007). "Identifying distributional characteristics in random coefficient panel data models," *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.

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

Behrman, Jere R. and Anil B. Deolalikar. (1987)."Will developing country nutrition improve with income? A case study for rural south India," *Journal of Political Economy* 95 (3): 492 - 507.

Behrman, Jere R. and Barbara L. Wolfe. (1983). "Is income overrated in determining adequate nutrition?" *Economic Development and Cultural Change* 31 (4): 525 - 549.

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

Bliss, Christopher and Nicholas Stern. (1978). "Productivity, wages and nutrition : Part I: the theory," *Journal of Development Economics* 5 (4): 331 - 362.

+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.

Bouis, Howarth E. (1994). "The effect of income on demand for food in poor countries: are our food consumption databases giving us reliable estimates? *Journal of Development Economics* 44(1): 199-226.

Bouis, Howarth E. and Lawrence J. Haddad. (1992)."Are estimates of calorie-income elasticities too high? A recalibration of the plausible range," *Journal of Development Economics* 39 (2): 333 - 364.

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.

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. (1995). "Earnings, schooling, and ability revisted," *Research in Labor Economics* 14: 23 - 48 (S. W. Polachek, Ed.). London: JAI Press, Ltd.

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..

Case, Anne. and Cristina Paxson. (2006). "Stature and status: height, ability, and labor market outcomes," *Mimeo*, Princeton University.

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

Chamberlain, Gary. (1980b). "Multivariate regression models for panel data," *NBER Technical Paper No. 8.*

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. (1992a). "Efficiency bounds for semiparametric regression," *Econometrica* 60 (3): 567 - 596.

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

Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn and Whitney Newey. (2007). “Identification and estimation of marginal effects in panel binary choice models with unobserved individual effects,” *Mimeo.*

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.

Dasgupta, Partha. (1993). *An Inquiry into Well-Being and Destitution.* Oxford: Oxford University Press.

Dasgupta, Partha and Debraj Ray. (1986). “Inequality as a determinant of malnutrition and unemployment:theory,” *Economic Journal* 96 (384): 1011 - 1034.

Dasgupta, Partha and Debraj Ray. (1987). “Inequality as a determinant of malnutrition and unemployment:policy,” *Economic Journal* 97 (385): 177 - 188.

Deaton, Angus and Salman Zaidi. (2002). “Guidelines for constructing consumption aggregates for welfare analysis,” *LSMS Working Paper Number 135.*

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.

Edmundson, W. C. and P. V. Sukhatme. (1990).“Food and work: poverty and hunger,” *Economic Development and Cultural Change* 38 (2): 263 - 280.

Fernández-Val, Iván. (2005). “Bias correction in panel data models with individual-specific parameters,” *Mimeo.*

Ferro-Luzzi, Anna. (2005). “The conceptual framework for estimating food energy requirement,”

Florens, Jean Pierre, James J. Heckman, Costas. Meghir and Edward Vytlacil. (2008). “Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogenous effects,” *Econometrica* 76 (5): 1191 - 1206.

Food and Agricultural Organization (FAO). (2001). "Human energy requirements: report of a joint FAO/WHO/UNU expert consultation," *FAO Food and Nutrition Technical Report Series 1.*

Food and Agricultural Organization (FAO). (2006). *The State of Food Insecurity in the World 2006.* Rome: Food and Agricultural Organization.

Graham, Bryan S., Jinyong Hahn and James L. Powell. (2008). "Quantile panel data models," *Mimeo.*

Garen, J. (1984). "The returns to schooling: a selectivity bias approach with a continuous choice variable," *Econometrica* 52 (5): 1199 - 1218.

Grantham-McGregor, Sally and Helen Baker-Henningham. (2005). "Review of the evidence linking protein and energy to mental development," *Public Health Nutrition* 8(7A): 1191 - 1201.

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 Economic Statistics* 19 (1): 16 - 17.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd. (1998). "Characterizing selection bias using experimental data," *Econometrica* 66 (5): 1017 - 1098.

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 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.

Heckman, James and Edward Vytlacil. (2001). "Local instrumental variables," *Nonlinear Statistical Modelling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*: 1 - 46 (C. Hsiao, K. Morimune, J. Powell, Eds.). Cambridge: Cambridge University Press.

Hoddinott, John, Emmanuel Skoufias and Ryan Washburn. (2000). "The impact of Progresa on consumption: a final report," *Mimeo.*

Hoddinott, John et al. (2008). "Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults," *Lancet* 371 (9610): 411 - 416.

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

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 Joshua D. Angrist. (1994). "Identification and estimation of local average treatment effects," *Econometrica* 62 (2): 467 - 475.

Imbens, Guido W. and Whitney K. Newey. (2007). "Identification and estimation of triangular simultaneous equations models without additivity," *Mimeo.*

Instituto Nacional de Estadísticas y Censos (INEC). (2005). *Metodología de construcción del agregado de consumo, de las líneas de pobreza y del agregado de ingreso en Nicaragua.* Retrieved January 22, 2008, from the World Wide Web: http://www.inec.gob.ni/Pobreza/pubmetodol.htm.

Instituto de Nutrición de Centro América y Panamá (INCAP) and Organización Panamericana la Salud (OPS). (2000). *Tabla de Composición de Alimentos de Centroamérica.* Retrieved January 2008, from the World Wide Web: http://www.tabladealimentos.net/tca/TablaAlimentos/inicio.html

International Food Policy Research Institute (IFPRI). (2005). *Nicaraguan RPS evaluation data (2000-02): overview and description of data files (April 2005 Release).* Washington D.C.: International Food Policy Research Institute.

Lewbel, Arthur. (1991). "The rank of demand systems: theory and nonparametric estimation," *Econometrica* 59 (3): 711 – 730.

Logon, Trevon. (2006). "Nutrition and well-being in the late nineteenth century," *Journal of Economic History* 66 (2): 313 - 341.

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

Manski, Charles F. (2003). *Partial Identification of Probability Distributions.* New York: Springer-Verlag.

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

Mirlees, James A. (1975). "A pure theory of underdeveloped economies," *Agriculture in Development Theory* (L.A. Reynolds, Eds.). New Haven: Yale University Press.

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

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

Mundlak, Yair. (1978b). "Models with variable coefficients: integration and extension," *Annales de l'insee* 30-31: 483 - 510.

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. and Daniel McFadden. (1994). "Large sample estimation and hypothesis testing," *Handbook of Econometrics 4*: 2111 - 2245 (R.F. Engle & D.L. McFadden). Amsterdam: North Holland.

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

Skoufias, Emmanuel. (2003). "Is the calorie-income elasticity sensitive to price changes? Evidence from Indonesia. *World Development* 31 (7): 1291 - 1307.

Smith, Lisa C. and Ali Subandoro. (2007). *Measuring food security using household expenditure surveys.* Washington D.C.:

International Food Policy Research Institute.

Stiglitz, Joseph E. (1976). "The efficiency wage hypothesis, surplus labor, and the distribution of income in L.D.C.s," *Oxford Economic Papers* 28 (2): 185 - 207.

Strauss, John and Duncan Thomas. (1995). "Human resources: empirical modeling of household and family decisions," *Handbook of Development Economics* 3 (1): 1883 - 2023. (J. Behrman & T.N. Srinivasan). Amsterdam: North-Holland.

Subramanian, Shankar and Angus Deaton. (1996). "The demand for food and calories," *Journal of Political Economy* 104 (1): 133 - 162.

Wolfe, Barbara L. and Jere R. Behrman. (1983). "Is income overrated in determining adequate nutrition?" *Economic Development and Cultural Change* 31 (3): 525 - 549.

Wooldridge. Jeffrey M. (1997). "On two stage least squares estimation of the average treatment effect in a random coefficient model," *Economic Letters* 56 (2): 129 - 133.

Wooldridge. Jeffrey M. (1999). "Distribution-free estimation of some nonlinear panel data models," *Journal of Econometrics* 90 (1): 77 - 97.

Wooldridge. Jeffrey M. (2003). "Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model," *Economic Letters* 79 (2): 185 - 191.

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.

World Bank. (2002). *Basic information document: Nicaragua living standards measurement survey 1998.* Washington D.C.: The World Bank.

World Bank. (2003). *Nicaragua Poverty Assessment: Raising Welfare and Reducing Vulnerability.* Washington D.C.: The World Bank.

A quantile panel data regression model[1]

Bryan S. Graham[†], Jinyong Hahn[♮], and James L. Powell[◇]

INITIAL DRAFT: March 2008
THIS DRAFT: November 4, 2008

PRELIMINARY AND INCOMPLETE
COMMENTS WELCOME

Abstract

To be completed....

JEL CLASSIFICATION: C14, C23, C33

KEY WORDS: PANEL DATA, CORRELATED RANDOM COEFFICIENTS, QUANTILE REGRESSION

---

[1] We would like to thank Manuel Arellano 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 - Los Angeles, Box 951477, Los Angeles, CA 90095-1477. E-MAIL: hahn@econ.ucla.edu. WEB: http://www.econ.ucla.edu/people/faculty/Hahn.html

[◇] 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

Quantile regression methods feature in empirical economic research with increasing regularity (e.g., Buchinsky, 1994; Mwabu and Schultz, 1996; Deaton, 1997; Arias, Hallock and Sosa-Escudero, 2001). At the same time recent theoretical research in econometrics substantially expands the range of problems to which such methods may be applied. Examples include Angrist, Chernozkukov and Fernández-Val's (2006) characterization of the properties of the linear quantile regression estimator of Koenker and Bassett (1978) under misspecification (cf., Chamberlain, 1994), Chernozhukov and Hansen's (2005, 2006), Ma and Koenker's (2006) and Imbens and Newey's (2008) alternative approaches to allowing for endogenous regressors, and Firpo's (2007) results on identification and estimation of quantile treatment effects under exogeneity (cf., Lehmann, 1974; Doksum, 1974). A comprehensive survey of quantile regression methods may be found in the monograph by Koenker (2005) (cf., Buchinsky, 1998; Koenker and Hallock, 2001).

Despite this recent surge of applied and theoretical work, the application of quantile regression methods to panel data remains relatively unexplored.[2] This is unfortunate since the availability of multiple observations of the same unit over time remains one of the principal ways of 'controlling for' unobserved correlated heterogeneity in applied work (e.g., Griliches, 1979; Currie and Thomas, 1995; Card, 1996; Altonji and Dunn, 1996).

In this paper we propose an extension of quantile regression methods to a variant of the correlated random coefficients panel data model analyzed by Chamberlain (1992), Graham and Powell (2008) and Arellano and Bonhomme (2008). In the two-period version of our model we let $\{Y_{1i}, Y_{2i}, X_{1i}, X_{2i}, A_i\}_{i=1}^{\infty}$ be an independently and identically distributed random sequence drawn from F_0 and maintain a potential, or structural, outcome equation of the form

$$\begin{aligned} Y_t(x_t) &= b_t(A, U_t) + c(A, U_t)\, x_t, \qquad t = 1, 2 \\ &= B_t + C_t x_t. \end{aligned} \tag{1}$$

The terms A and U_t respectively correspond to time-invariant and time-varying unobserved heterogeneity.[3] The functions $b_t(A, U_t)$ and $c(A, U_t)$ map this heterogeneity into the period-by-unit-specific intercept and slope coefficients, respectively $B_t \overset{def}{\equiv} b_t(A, U_t)$ and $C_t \overset{def}{\equiv} c(A, U_t)$. Our motivation for analyzing panel data is to allow (B_t, C_t) to be dependent on X_t; that is to allow for X_t to be 'endogenous'. We view (1) as structural in the sense that $Y_t(x)$ traces out a unit's period-t outcome across different hypothetical values of x_t. The observed outcome corresponds to $Y_t = Y_t(X_t)$.

The unit-specific effect of a small increase in X_t is given by

$$\frac{\partial Y_t(x_t)}{\partial x_t} = C_t = c(A, U_t),$$

[2] Neither of the *Handbook of Econometrics* chapters on panel data discuss extensions to quantile regression (Chamberlain, 1984, Arellano and Honoré, 2001).

[3] In what follows we use capital letters to denote random variables, lower case letters specification realizations of them, and calligraphic letters their support.

which, due to linearity, does not depend on x_t. In Graham and Powell (2008) we discuss conditions ensuring identification of the average partial effect (APE) – $\gamma = \mathbb{E}[C_t]$ – and propose a consistent estimator. In this paper we consider identification and estimation of the *quantile partial effect* (QPE), $\gamma_t(\tau) = F_{C_t}^{-1}(\tau) = \inf\{c \in \mathbb{R} : F_{C_t}(\tau) \geq \tau\}$; $\gamma_t(\tau)$ equals the τ^{th} quantile of C_t. If X_t is exogenously increased by one unit for all individuals, then 100τ percent of the population will experience a change in outcomes less than $\gamma_t(\tau)$, with the remainder experiencing a larger outcome effect. Knowledge of the QPE allows for a complete characterization of the marginal effect of changes in X_t on the distribution of outcomes.

Koenker and Hallock (2000) speculate the one reason for the limited progress on extending quantile regression techniques to panel data, may be that identification strategies based on 'differencing away' unobserved correlated heterogeneity appear difficult to apply. These difficulties stem from the fact that conditional quantiles, unlike conditional expectations, are not linear operators (cf., Canay, 2008). Nevertheless we show that a natural generalization of the textbook linear quantile regression model of Koenker and Bassett (1978) (henceforth the QR model) is appropriate for the analysis of panel data.

Our extension begins from the observation that the linear QR model has a random coefficient representation (e.g., Koenker, 2005, pp. 59 - 62). These random coefficients are (i) independent of the regressor and (ii) comonotonic or perfectly rank correlated (perhaps after reparameterization). Neither of these two assumptions is particularly attractive and we show how the availability of panel data allows one to weaken them. In particular, panel data allow us to (i) introduce relatively arbitrary forms of dependence between the random coefficients and all leads and lags of the regressors, (ii) weaken the comonotonicity requirement such that it only needs to hold within subpopulations with common histories of regressor values and (iii) allow for a reasonably rich specification of aggregate time effects.

Our basic setup, while restrictive, is consistent with the intuition that panel data allow the researcher to work with weaker assumptions than those required for cross section analyses. Importantly our framework encompasses standard panel data models as special cases. For example, if $b_t(A, U_t) = \delta_t + A + U_t$ and $c(A, U_t) = \gamma$ we have the textbook linear panel data model (cf., Mundlak, 1961; Chamberlain, 1982, 1984):

$$Y_t = \delta_t + \gamma X_t + A + U_t. \tag{2}$$

Chamberlain (1982, 1984) discusses identification and estimation of this model under the strict exogeneity restriction $\mathbb{E}[U_t | X, A] = 0$ with $X = (X_1, X_2)'$.

If, instead, $b_t(A, U_t) = \delta_t(A + U_t)$ we get a model with time-varying loadings on the unobserved heterogeneity

$$Y_t = \gamma X_t + \delta_t A + U_t^*, \tag{3}$$

with $U_t^* = \delta_t U_t$. This model is also considered by Chamberlain (1982, 1984). The case where Y_t is censored is analyzed by Honoré (1992), Chay and Honoré (1992) and Chen and Khan (2008).

Unlike the above examples, we are interested in settings where the distribution of both $B_t = b_t(A, U_t)$ and $C_t = c(A, U_t)$ are non-degenerate and possibly dependent on X. Such dependence renders the application of standard quantile regression methods inappropriate. Our contribution, therefore, is to provide assumptions under which panel data allows for identification and consistent estimation of $\gamma_t(\tau)$ under such dependence.

Koenker (2004) considers an application of quantile regression methods to longitudinal data where the unit-specific intercepts are pure location-shifts and hence invariant across conditional quantiles. Canay (2008), motivated by incidental parameter concerns, considers the same model in a large-T setting. Abrevaya and Dahl (2008) work instead with a heuristic adaptation of Chamberlain's (1982, 1984) correlated random effects (CRE) structure. In an interesting application of quantile regression methods to panel data, Gamper-Rabindran, Khan and Timmins (2008), also place a CRE structure on a unit-specific intercept. They show that the resulting model may be estimated using the censored panel data regression estimator of Chen and Khan (2008).

Our work differs from each of these papers, most importantly, in allowing correlated random slope coefficients in addition to unit-specific intercepts. Abrevaya and Dahl (2008) target estimand differs in general from the QPE, except, as we show formally below, for the special case where C_t is independent of X_t. From this perspective our work both provides a explicit generative model justifying Abrevaya and Dahl's (2008) target estimand and generalizes to the case of correlated slope heterogeneity. A similar connection exists between the procedure used in Gamper-Rabindran, Khan and Timmins (2008) and our own. We detail these connections in Section 2 below.

The next section reviews the linear QR model for cross section data and outlines our extension of it to panel data. In this section we work with a basic two period, scalar regressor, formulation of our model. Section 3 presents our approach to estimation. In Section 4 we discuss a variety of extensions of our basic model, including one appropriate for the multi-period, multi-regressor setting. Section 5 summarizes and suggests areas for future research.

2 Identification

2.1 The linear quantile regression model

It is useful to begin with a review of the textbook linear quantile regression model for cross section data. As noted in the introduction, this model has a random coefficients or Skorohod representation of (e.g., Koenker, 2005, pp. 59 - 62)

$$Y(x) = b(U) + c(U)x, \quad U|X \sim \mathcal{U}[0,1].$$

Let $Q_Y(\tau|X = x) = \inf\{y \in \mathbb{R} : F_Y(\tau|x) \geq \tau\}$ denote the conditional quantile function (CQF). To ensure monotonicity of $Q_Y(\tau|X = x)$ in τ we also require that both $b(U)$ and $c(U)$ are non-decreasing in U and $x \in \mathbb{R}^+$. Under these assumptions

$$Q_Y(\tau|X = x) = \beta(\tau) + \gamma(\tau)x,$$

with $\beta(\tau) = F_B^{-1}(\tau) = Q_B(\tau)$ and $\gamma(\tau) = F_C^{-1}(\tau) = Q_C(\tau)$.

Let $B = b(U)$ and $C = c(U)$, an alternative representation of the model, which emphasizes independence and co-monotonicity of the random coefficients, is

$$Y(x) = B + Cx, \quad x \in \mathbb{R}^+ \tag{4}$$

with the conditional distribution of (B, C) given X equal to

$$(B, C)|\, X \stackrel{D}{=} \left(F_B^{-1}(V), F_C^{-1}(V)\right), \qquad V \sim \mathcal{U}[0, 1]. \tag{5}$$

As (4) and (5) highlight, linearity of the conditional quantile function requires the imposition of (i) independence of the random coefficients and the regressor and (ii) of a strong form of dependence across the random coefficients. In particular, (5) implies that for any set of pairs (B, C) and (B', C') such that $B \geq B'$, then $C \geq C'$ (i.e., B and C have a rank correlation of one). Chernozhukov and Hansen (2005, 2006) refer to a related restriction as rank invariance. Mathematically this property of the joint distribution of B and C is known as comonotonicity (cf., Jouini and Napp, 2004).

Both independence and comonotonicity are unattractive restrictions (cf., Heckman, Smith and Clements, 1997). Recent work on nonseparable models with endogenous regressors emphasizes that the economics of optimization often implies that the marginal return to a unit increase in an input will be correlated with the chosen level of that input (e.g., Imbens, 2007). For example, an individual's return to an additional year of school will generally be correlated with her chosen number of years of completed schooling (e.g., Card, 2001).

Comonotonicity also may be difficult to motivate, at least in some settings. In the context of earnings-schooling models it requires that those individuals with the greatest 'absolute advantage' in the labor market (i.e., high B individuals), have the greatest 'comparative advantage' in schooling (i.e., are also high C individuals). In practice we might believe that those individuals who would earn high wages with zero years of schooling, would earn relatively lower wages in occupations requiring many years of schooling (and vice-versa).

2.2 A panel data extension

Panel data allow us to relax the assumptions of both independence and comonotonicity. We begin with the simple two-period case under random sampling. The simplest version of the model we consider is defined by the following set of assumptions.

Assumption 2.1 $\{Y_{1i}, Y_{2i}, X_{1i}, X_{2i}, A_i\}_{i=1}^{\infty}$ *be an independently and identically distributed random sequence drawn from* F_0.

Assumption 2.1 ensures that the joint distributions of $(Y', X')'$, with $X = (X_1, X_2)'$ and $Y = (Y_1, Y_2)'$, is asymptotically identified.

Let A be time-invariant, unit-specific, unobserved heterogeneity and U_t time-varying heterogeneity. We assume that outcomes are generated by a correlated random coefficients model.

Assumption 2.2 *The outcome equation is of the form*

$$\begin{aligned} Y_t &= b_t(A, U_t) + c(A, U_t) X_t, \qquad t = 1, 2 \\ &= B_t + C_t X_t, \end{aligned}$$

with $B_t = b_t(A, U_t)$ *and* $C_t = c(A, U_t)$.

Note that the function mapping A and U_t into the unit-specific slope coefficient – $c(A, U_t)$ – is assumed constant across the two periods; that is $c(A, U_1) = c(A, U_2)$ if $U_1 = U_2$. However, we do allow the function mapping A and U_t into the unit-specific intercept – $b_t(A, U_t)$ – to vary over time (albeit in a way that is restricted by Assumption 2.5 below).

As in Manski (1987) and Honoré (1992) we assume that conditional on X and A the distribution of U_t is stationary over time.

Assumption 2.3 (Marginal Stationarity) *(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* $(X, A) \in \mathcal{X} \times \mathcal{A}$.

Assumption 2.3 does not restrict the conditional distribution of A given X. In this sense, as in Graham and Powell (2008), we maintain a 'fixed effect' characterization of A. Nevertheless Assumption 2.3, while allowing for serial dependence in U_t and certain forms of heteroscedasticity, is restrictive.[4]

Our next assumption relaxes, relative to the cross section QR model, the requirement of intercept and slope comonotonicity.

Assumption 2.4 (Conditional Comonotonicity) *(i)* $b_t(A, U_t) = b_t^*(V_t, X)$ *and* $c(A, U_t) = c^*(V_t, X)$ *for some scalar random variable* V_t; *(ii)* $b_t^*(v_t, x)$ *and* $c^*(v_t, x)$ *are non-decreasing in* v_t; *(iii)* $X_t \in \mathcal{X}_t \subset \mathbb{R}^+$.

An alternative characterization of Assumption 2.4 is as follow. Recalling that $B_t = b_t(A, U_t)$ and $C_t = c(A, U_t)$, Assumption 2.4 implies that we may represent the period t joint distribution of (B_t, C_t) conditional on $X = (X_1, X_2)'$ as

$$(B_t, C_t)| X \overset{D}{=} \left(F_{B_t}^{-1}(V | X_1, \ldots, X_T), F_{C_t}^{-1}(V | X_1, \ldots, X_T)\right), \qquad V \sim \mathcal{U}[0, 1]. \tag{6}$$

As in the model for cross section data we impose a strong form of dependence across the random coefficients. However, the availability of panel data, allow us to relax two features of the cross section model. First we can allow the joint distribution of $(B_t, C_t)| X$ to depend on X. Second, we only require that co-monotonicity of B_t and C_t hold within subpopulations homogenous in X. The

[4] For example it rules out heteroscedasticity over time (cf., Arellano, 2003).

panel allows us to introduce correlated heterogeneity *and* weaken the dependence structure across the random coefficients.

Consider the estimation of the union wage premium as in Chamberlain (1994) and Card (1996) amongst others. Let X_t be a binary indicator for coverage by a collective bargaining agreement in period t. Unconditional comonotonicity requires perfect rank correlation across the distribution of (potential) earnings in, respectively, the absence and presence of collective bargaining coverage (i.e., B_t and $B_t + C_t$). In contrast Assumption 2.4 only requires such perfect rank correlation within subpopulations with identical coverage histories (e.g., among those individuals always covered by collective bargaining). While still a rather strong assumption, it is evidently weaker that its unconditional counterpart.

Assumption 2.4 restricts the ways in which A and U_t affect the intercept and slope coefficients. To see the types of restrictions involved it is helpful to consider a specific example. Let $b_t(A, U_t) = \widetilde{b}_t(A + U_t)$ with $\widetilde{b}_t(\cdot)$ increasing in its argument and similarly for $c(A, U_t) = \widetilde{c}(A + U_t)$. Let $V|X = x \sim \mathcal{U}[0, 1]$, then the random variable $F^{-1}_{A+U_t|X}(V|X)$ has distribution $F_{A+U_t|X}$, allowing us to write $\widetilde{b}_t(A + U_t) = \widetilde{b}_t(F^{-1}_{A+U_t|X}(V|X)) = b^*_t(V_t, X)$ and similarly for $\widetilde{c}(A + U_t)$. This representation satisfies the requirements of conditional comonotonicity. A key requirement of Assumption 2.4 is the ability to represent the coefficients in terms of a single underlying random variable. This, in turns, requires that A and U_t enter $b_t(A, U_t)$ and $c(A, U_t)$ in particular ways.

As noted above we allow for the function $b_t(a, u_t)$ to vary over time. Our next assumption restricts the allowable form of such variation.

Assumption 2.5 (Additive Quantile Trends) *For all $\tau \in (0, 1)$ and $x \in \mathcal{X}$*

$$Q_{B_2|X}(\tau|x) = Q_{B_1|X}(\tau|x) + \delta(\tau).$$

Assumption 2.5 restricts the 'horizontal distance' between $F_{B_1|X}$ and $F_{B_2|X}$ at b to not depend on x. This distance, $\Delta(b)$, is implicitly defined by

$$F_{B_1|X}(b|X) = F_{B_2|X}(b + \Delta(b)|X).$$

Solving for $\Delta(b)$ yields

$$\Delta(b) = F^{-1}_{B_2|X}\left(F_{B_1|X}(b|X)\big|X\right) - b,$$

which after changing variables to $\tau = F_{B_1|X}(b|X)$ gives $Q_{B_2|X}(\tau|x) = Q_{B_1|X}(\tau|x) + \delta(\tau)$. That is, Assumption 2.5 implies that the τ^{th} conditional quantile of B_1 additively shifts by δ_τ across all subpopulations defined in terms of X. While Assumption 2.5 is more restrictive than we require (we explore weaker forms of 'common trend' assumptions in Section 4 below), it is nevertheless weaker than imposing a traditional additive-shift form of time effects. In the current setting this would correspond to the assumption that $\Delta(b)$ is constant in b.

An alternative to Assumption 2.5 is:

Assumption 2.6 (Multiplicative Quantile Trends) *For all* $\tau \in (0,1)$ *and* $x \in \mathcal{X}$

$$Q_{B_2|X}(\tau|\, x) = \delta(\tau)\, Q_{B_1|X}(\tau|\, x).$$

Assumption 2.5 implies that the horizontal distance between $F_{B_1|X}$ and $F_{B_2|X}$ at b is multiplicative in b and constant in x :

$$F_{B_1|X}(b|\, X) = F_{B_2|X}(\Delta(b)\, b|\, X),$$

which, after solving for $\Delta(b)\, b$ and changing variables to $\tau = F_{B_1|X}(b|\, X)$, yields Assumption 2.5 or the requirement that the τ^{th} conditional quantile of B_2 is a multiple of the corresponding B_1 quantile.

The time-varying factor loading models studied by Honoré (1992), Chay and Honoré (1998) and Chen and Khan (2008) correspond to the case where $\delta(\tau) = \delta$ for all $\tau \in (0,1)$.

Our final two assumptions are more technical in nature. First we require smoothness in the conditional quantile functions of the random slope coefficients.

Assumption 2.7 (Smoothness) $Q_{C_1|X}(\tau|\, x)$ *and* $Q_{C_2|X}(\tau|\, x)$ *are continuous and differentiable in* x.

Second we require that the support of $X = (X_1, X_2)'$ includes realizations of X corresponding to 'stayers' or units who do not switch their value of X_t between periods one and two.

Assumption 2.8 (Presence of Stayers)

$$\mathcal{S} = \{x : f_X(x) > 0,\ x_1 = x_2\} \neq \varnothing.$$

Assumption 2.8 suggests that identification may be fragile in settings where X_t has a strong deterministic drift component.

Our first result is to show that unconditional quantiles of C_t are identified.

Theorem 2.1 *Under Assumptions 2.1 to 2.5 and 2.7 and 2.8 the* τ^{th} *quantile of* C_t *is identified by the solution to*

$$\mathbb{E}\left[\int_{v=0}^{v=1} \mathbf{1}(\gamma(v;x) \leq \gamma(\tau))\, \mathrm{d}v - \tau\right] = 0, \tag{7}$$

with $\gamma(\tau; x)$ *given by*

$$\gamma(\tau; x) = \frac{Q_{Y_2|X}(\tau|\, x) - Q_{Y_1|X}(\tau|\, x) - \left\{Q_{Y_2|X}(\tau|\, x^*) - Q_{Y_1|X}(\tau|\, x^*)\right\}}{x_2 - x_1} \tag{8}$$

for $x \in \mathcal{S}^c$ *(i.e., for movers) and* x^* *some element of* $\mathcal{S}$*, and by*

$$\gamma(\tau; x) = \lim_{h \downarrow 0} \frac{Q_{Y_2|X}(\tau|\, x_1, x_1 + h) - Q_{Y_1|X}(\tau|\, x) - \left\{Q_{Y_2|X}(\tau|\, x^*) - Q_{Y_1|X}(\tau|\, x^*)\right\}}{h}. \tag{9}$$

for $x \in \mathcal{S}$ (i.e., for stayers).

Proof. The structure of our result hinges on the restricted forms for $Q_{Y_1|X}(\tau|x)$ and $Q_{Y_2|X}(\tau|x)$ implied by the model. This ensures that the conditional quantiles of C_t are identified. We then recover the unconditional quantiles in a manner similar to that of Gosling, Machin and Meghir (2000) and Melly (2005, 2007) (cf., Machado and Mata, 2005). Under Assumptions 2.2 and 2.4 we have

$$\begin{aligned}
Q_{Y_1|X}(\tau|x) &= Q_{B_1|X}(\tau|x) + Q_{C_1|X}(\tau|x)x_1 \\
Q_{Y_2|X}(\tau|x) &= Q_{B_2|X}(\tau|x) + Q_{C_2|X}(\tau|x)x_2.
\end{aligned}$$

Assumptions 2.3 and 2.5 imply that $Q_{C_1|X}(\tau|x) = Q_{C_2|X}(\tau|x) \overset{def}{\equiv} \gamma(\tau;x)$ and $Q_{B_2|X}(\tau|x) = \beta(\tau;x) + \delta(\tau)$ with $\beta(\tau;x) \overset{def}{\equiv} Q_{B_1|X}(\tau|x)$. This yields the simplification

$$\begin{aligned}
Q_{Y_1|X}(\tau|x) &= \beta(\tau;x) + \gamma(\tau;x)x_1 \\
Q_{Y_2|X}(\tau|x) &= \delta(\tau) + \beta(\tau;x) + \gamma(\tau;x)x_2.
\end{aligned}$$

For $x^* \in \mathcal{S}$ we have

$$Q_{Y_2|X}(\tau|x^*) - Q_{Y_1|X}(\tau|x^*) = \delta(\tau), \tag{10}$$

and hence for $x \in \mathcal{S}^c$ we have (8) in the Theorem. Under Assumption 2.7 $\gamma(\tau;x)$ is identified by the limit (9) for $x \in \mathcal{S}$.

With $\gamma(\tau;x)$ identified for all $x \in \mathcal{X}$ and $F_X(x)$ asymptotically revealed by the sampling process (Assumption 2.1) we may recover the quantile partial effects (QPEs) $\gamma(\tau) = Q_{C_1}(\tau) = F_{C_1}^{-1}(\tau)$ as follows. Let $V|X = x \sim \mathcal{U}[0,1]$, then the random variable $F_{C_1|X}^{-1}(V|X)$ has distribution $F_{C_1|X}$. Using this fact, iterated expectations and the change-of-variables formula we have $\gamma(\tau)$ identified by the solution to

$$\begin{aligned}
\tau &= \int \mathbf{1}(c \leq \gamma(\tau)) f_{C_1}(c)\,\mathrm{d}c \\
&= \int \left\{ \int \mathbf{1}(c \leq \gamma(\tau)) f_{C_1|X}(c|x)\,\mathrm{d}c \right\} f_X(x)\,\mathrm{d}x \\
&= \int \left\{ \int_{v=0}^{v=1} \mathbf{1}(\gamma(v;x) \leq \gamma(\tau))\,\mathrm{d}v \right\} f_X(x)\,\mathrm{d}x,
\end{aligned}$$

as claimed. ■

Note that if instead of additive trends, we assume multiplicative, trends (i.e., Assumption 2.6 instead of 2.5), (7) still identifies $\gamma(\tau)$. However (10) no longer identifies $\delta(\tau)$, which is instead given by $\delta(\tau) = Q_{Y_2|X}(\tau|x^*)/Q_{Y_1|X}(\tau|x^*)$ for some $x^* \in \mathcal{S}$.

2.3 Relationship to prior research

Consider the textbook linear panel data model

$$Y_t = \delta_t + \gamma X_t + A + U_t, \qquad \mathbb{E}\left[U_t \middle| X, A\right] = 0.$$

Chamberlain (1982, 1984) noted that under this model

$$\gamma = \frac{\partial \mathbb{E}^*\left[Y_2 \middle| X\right]}{\partial X_2} - \frac{\partial \mathbb{E}^*\left[Y_1 \middle| X\right]}{\partial X_2},$$

where $\mathbb{E}^*\left[Y_2 \middle| X\right]$ denotes the mean squared error minimizing linear predictor (LP) of Y_2 given X.

Abrevaya and Dahl (2008) use this observation to heuristically motivate a definition of the quantile partial effect equal to

$$\gamma^*(\tau) = \frac{\partial Q_{Y_2|X}\left(\tau \middle| x\right)}{\partial x_2} - \frac{\partial Q_{Y_1|X}\left(\tau \middle| x\right)}{\partial x_2}.$$

In the setup considered here we have

$$\begin{aligned}
\frac{\partial Q_{Y_1|X}\left(\tau \middle| x\right)}{\partial x_1} &= \frac{\partial \beta(\tau; x)}{\partial x_2} + \frac{\partial \gamma(\tau; x)}{\partial x_2} x_1 \\
\frac{\partial Q_{Y_2|X}\left(\tau \middle| x\right)}{\partial x_2} &= \frac{\partial \beta(\tau; x)}{\partial x_2} + \frac{\partial \gamma(\tau; x)}{\partial x_2} x_2 + \gamma(\tau; x),
\end{aligned}$$

and hence Abrevaya and Dahl's (2008) estimand, $\gamma^*(\tau)$, is equal to

$$\frac{\partial Q_{Y_2|X}\left(\tau \middle| x\right)}{\partial x_2} - \frac{\partial Q_{Y_1|X}\left(\tau \middle| x\right)}{\partial x_2} = \gamma(\tau; x) + \left\{ \frac{\partial \gamma(\tau; x)}{\partial x_2} x_2 - \frac{\partial \gamma(\tau; x)}{\partial x_2} x_1 \right\}, \tag{11}$$

which differs, in general, from $\gamma(\tau)$, the QPE.

However if we augment our model with the additional restriction that $\gamma(\tau; x)$ is constant in x, then (11) does indeed identify the QPE. This restriction would follow from independence of the random slope coefficient and the regressors.

3 Estimation

4 Extensions

5 Conclusion

References

[1] Abrevaya, Jason and Christian M. Dahl. (2008). "The effects of birth inputs on birthweight: evidence from quantile estimation on panel data," *Journal of Business and Economic Statistics*

26 (4): 379 - 397.

[2] Altonji, Joseph G. and Thomas A. Dunn. (1996). "Using siblings to estimate the effect of school quality on wages," *Review of Economics and Statistics* 78 (4): 665 - 671.

[3] Angrist, Joshua, Victor Chernozhukov, and Iván Fernández-Val. (2006). "Quantile regression under misspecification, with an application to the U.S. Wage Structure," *Econometrica* 74 (2): 539 - 563.

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

[5] 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.

[6] Arias, Omar, Kevin F. Hallock and Walter Sosa-Escudero. (2001). "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data," *Empirical Economics* 26 (1): 7 - 40.

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

[8] Blundell, Richard W. and James L. Powell. (2003). "Endogeneity in nonparametric and semi-parametric 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.

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

[10] 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.

[11] Buchinsky, Moshe. (1994). "Changes in the U.S. wage structure 1963-1987: application of quantile regression," *Econometrica* 62 (2): 405 - 458.

[12] Buchinsky, Moshe. (1998). "Recent advances in quantile regression models: a practical guideline for empirical research," *Journal of Human Resources* 33 (1): 88 - 126.

[13] Canay, Ivan A. (2008). "A note on quantile regression for panel data models," *Mimeo.*

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

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

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

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

[18] Chamberlain, Gary. (1992). "Efficiency bounds for semiparametric regression," *Econometrica* 60 (3): 567 - 596.

[19] Chamberlain, Gary. (1994). "Quantile regression, censoring, and the structure of wages," *Advances in Econometrics, Sixth World Congress* I: 171 - 209. (C. Sims, Ed.). Cambridge: Cambridge University Press.

[20] Chaudhuri, Probal. (1991). "Nonparametric estimates of regression quantiles and their local Bahadur representation," *Annals of Statistics* 19 (2): 760 - 777.

[21] Chay, Kenneth Y. and Bo E. Honoré. (1998). "Estimation of semiparametric censored regression models: an application to changes in Black-White earnings inequality during the 1960s," *Journal of Human Resources* 33 (1): 4 - 38.

[22] Chay, Kenneth Y. and James L. Powell. (2001). "Semiparametric censored regression models," *Journal of Economic Perspectives* 15 (4): 29 -42.

[23] Chen, Songnian and Shakeeb Khan. (2008). "Semiparametric estimation of nonstationary censored panel data models with time varying factor loads," *Econometric Theory* 24 (5): 1149 - 1173.

[24] Chen, Songnian, Gordon B. Dahl and Shakeeb Khan. (2005). "Nonparametric identification and estimation of a censored location-scale regression model," *Journal of the American Statistical Association* 100 (469): 212 - 221.

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

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

[27] Chernozhukov, Victor, Iván Fernández-Val and Blaise Melly. (2008). "Inference on counterfactual distributions," *Mimeo.*

[28] Chesher, Andrew. (2003). "Identification in nonseparable models," *Econometrica* 71 (5): 1405 - 1441.

[29] 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.

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

[31] Deaton, Angus. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore: The John Hopkins University Press.

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

[33] Firpo, Sergio. (2007). "Efficient semiparametric estimation of quantile treatment effects," *Econometrica* 75 (1): 259 - 276.

[34] Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. (2007). "Unconditional quantile regression," *NBER Technical Working Paper 339.*

[35] Gamper-Rabindran, Shanti, Shakeeb Khan and Christopher Timmins. (2008). "The impact of piped water provision on infant mortality in Brazil: a quantile panel data approach," *NBER Working Paper 14365.*

[36] Geraci, Marco and Matteo Bottai. (2007). "Quantile regression for longitudinal data using the asymmetric Laplace distribution," Biostatistics 8 (1): 140 - 154.

[37] Gosling, Amanda, Stephen Machin and Costas Meghir. (2000). "The changing distribution of male wages in the U.K.," *Review of Economic Studies* 67 (4): 635 - 666.

[38] Graham, Bryan S. and James L. Powell. (2008). "Identification and estimation of 'irregular' correlated random coefficient models," *NBER Working Paper No. 14469.*

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

[40] Harding, Matthew and Carlos Lamarche. (2008). "A quantile regression approach for estimating panel data models using instrumental variables," *Mimeo.*

[41] 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.

[42] Holland, Paul W. (2002). "Two measures of change in the gaps between the CDFs of test-score distributions," *Journal of Educational and Behavioral Statistics* 27 (3): 3 - 17.

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

[44] 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.

[45] Imbens, Guido W. and Whitney K. Newey. (2008). “Identification and estimation of triangular simultaneous equations models without additivity,” *Mimeo.*

[46] Jouini, Elyés and Clotilde Napp. (2004). “Conditional comonotonicity,” *Decisions in Economics and Finance* 27 (2): 153 - 166.

[47] Koenker, Roger. (2004). “Quantile regression for longitudinal data,” *Journal of Multivariate Analysis* 91 (1): 74 - 89.

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

[49] Koenker, Roger and Gilbert Bassett, Jr. (1978). “Regression quantiles,” *Econometrica* 46 (1): 33 - 50.

[50] Koenker, Roger and Kevin Hallock. (2000). “Quantile regression: an introduction,” Mimeo.

[51] Koenker, Roger and Kevin Hallock. (2001). “Quantile regression: an introduction,” *Journal of Economic Perspectives* 15 (4): 143 – 156.

[52] Ma, Lingjie and Roger Koenker. (2006). “Quantile regression methods for recursive structural equation models,” *Journal of Econometrics* 134 (2): 471 - 506.

[53] Machado, José A. F. and José Mata. (2005). “Counterfactual decomposition of changes in wage distributions using quantile regression,” *Journal of Applied Econometrics* 20 (4): 445 - 465.

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

[55] Melly, Blaise. (2005). “Decomposition of differences in distribution using quantile regression,” *Labour Economics* 12 (4): 577 - 590.

[56] Melly, Blaise. (2007). “Estimation of counterfactual distributions using quantile regression,” *Mimeo.*

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

[58] Mwabu, Germano and T. Paul Schultz. (1996). “Education returns across quantiles of the wage function: alternative explanations for returns to education in South Africa,” *American Economic Review* 86 (2): 335 - 339.

[59] Wooldridge, Jeffrey M. (2005). "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.