

Identification and estimation of ‘irregular’ correlated random coefficient models¹

Bryan S. Graham[†] and James L. Powell[◇]

INITIAL DRAFT: December 2007

THIS DRAFT: July 3, 2008

PRELIMINARY AND INCOMPLETE

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 average effects, over either the full population distribution of unobserved heterogeneity (the average partial effect, APE), or over subpopulations defined by their regressor values (local average responses, LARs). In our primary 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. Our model is a special case of Chamberlain’s (1992a) correlated random coefficients (CRC) model, but does not satisfy the regularity conditions he imposes. We show how two measures of the outcome and regressor for each unit are sufficient for identification of the APE and LAR, as well as aggregate trends that are mean-independent of the regressors. We identify aggregate trends using units with a zero first difference in the regressor, 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).

JEL CLASSIFICATION: C14, C23, C33

KEY WORDS: PANEL DATA, CORRELATED RANDOM COEFFICIENTS, CONTROL FUNCTION, AVERAGE PARTIAL EFFECTS, LOCAL AVERAGE RESPONSE, KERNEL REGRESSION, CALORIE DEMAND

¹We 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) and the JAE Conference on Distributional Dynamics (CEMFI, Madrid) for comments and feedback. Discussions with Manuel Arellano, Stéphane Bonhomme, Gary Chamberlain, Michael Jansson and Edward Vytlacil were helpful in numerous ways. Max Kasy provided excellent 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 $t = 1, \dots, T$ periods for N randomly sampled units. In the most basic model we analyze the structural outcome equation is given by

$$Y_t = a(A, U_t) + b(A, U_t) X_t \quad (1)$$

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

$$Y_t(x_t) = a(A, U_t) + b(A, U_t) x_t \quad (2)$$

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.

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) and the *local average response* (LAR) (cf., Altonji and Matzkin, 2005; Bester and Hansen, 2007). In the binary regressor case these two objects correspond to the average treatment effect (ATE) and the average treatment effect on the treated (ATT) (cf., Florens, Heckman, Meghir and Vytlačil, 2008).⁴

The average partial effect is given by

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

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

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

⁴In a companion paper we study quantile partial effects (QPEs) (Graham, Hahn and Powell, 2008).

Because of linearity of (2), β_t does not depend on x_t .⁵

Let $g(x)$ be some known function of $x = (x_1, \dots, x_T)'$, the local average response gives the average partial effect within a subpopulation defined by its choice of $g(X) = g(x)$ is given by

$$\gamma_t(g(x)) \equiv \mathbb{E} \left[\frac{\partial Y_t(x_t)}{\partial x_t} \Big| g(X) = g(x) \right] = \mathbb{E} [b(A, U_t) | g(X) = g(x)]. \quad (4)$$

If, for example, $g(x) = \iota'x/T$ (with ι a $T \times 1$ vector of ones), then $\gamma_t(g(x))$ gives the average effect of a unit change in x_t within subpopulations defined by their *average* choice of X_t . Identification and estimation of (3) and (4) is nontrivial because X_t may vary systematically with A and/or U_t . To see the consequences of this dependence observe that the derivative of the regression function of Y_t given X_t does not identify a structural parameter. Differentiating through the integral we have

$$\frac{\partial \mathbb{E}[Y_t | X_t = x_t]}{\partial x_t} = \gamma_t(x_t) + \int \int \{a(a, u_t) + b(a, u_t)x_t\} \frac{\partial f(a, u_t | x_t)}{\partial x_t} dm(a) dm(u_t).$$

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

To contextualize our contribution 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 and LAR 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 $U = (U_1, \dots, U_T)'$ and $X = (X_1, \dots, X_T)'$. 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).

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

⁵We also study the case where X_t is itself a function of a lower-dimensional choice variable R_t . In that case, the APE, defined in terms of r_t , may vary with r_t . Extending our results to this case is straightforward.

Manski (1987), Honoré (1992) and Abrevaya (2000), however we leave $F(A|X)$ nonparametric. In our setup both the APE and LAR are identified when X_t is continuously-valued. In fact, we are able to provide a characterization of when these estimands are semiparametrically just-identified. In that sense, our maintained assumptions are minimally sufficient (although not necessary).⁶

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).⁷

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).⁸ 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 a method-of-moments estimator for the APE, however, the regularity conditions required for his estimator, as we discuss further below, rule out substantively important economic models. 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 (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.

Chernozhukov, Fernández-Val, Hahn and Newey (2007) study identification of the APE in a probit model with unit-specific intercepts (in the index function) and a discrete or continuous scalar regressor. They show that the maximum likelihood estimator which estimates the unit-specific intercepts along with the coefficient on X_t can be used to construct bounds on the ATE despite the incidental parameters problem (cf., Hahn, 2001). Porter (1996) and Das (2003) study nonparametric estimation of panel data model with additive unobserved heterogeneity. Honoré (1992) and Abrevaya (2000) consider models with nonseparable heterogeneity but, like Manski (1987), only identify index coefficients, not the APE or LAR. Arellano and Bonhomme (2007) also

⁶Chesher (2007) provides an extended discussion of the value of ‘just identifying’ semiparametric restrictions.

⁷Much of this research is surveyed by Imbens (2007).

⁸The 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.”

study identification in Chamberlain’s (1992a) CRC model. Unlike us, they maintain his regularity conditions.

The next section reports identification results for the APE and LAR in a two period version of our core model. When X_t is discretely-valued our assumptions generally only bound the APE and LAR (appropriately defined to account for the discreteness of X_t). Our analysis of this case also suggests 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 APE and LAR are point identified. We also contrast our approach to identification with the semiparametric random effects methods employed 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. We explain the sources of these differences in detail.

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 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 almost forty percent of households have less than 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

⁹We 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.

¹⁰Another 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).

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.

Section 5.4 summarizes 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)

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

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

Assumption 2.2 (MARGINAL STATIONARITY) (i)

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

(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

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

denote the average effect of a small change in x_t within the subpopulation of units where $X = x = (x_1, x_2)'$. Observe that $\beta_t(x)$, while closely related, is distinct from the LAR. It gives the average effect within a subpopulation defined by a common complete *history* of choices for X_t . Our first result shows that $\beta_t(x)$ is just-identified when $x_1 \neq x_2$.

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

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

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

$$\beta_t(X) = \mathbb{E}[b(A, U_t) | X] = \mathbb{E}[\mathbb{E}[b(A, U_t) | X, A] | X] = \mathbb{E}[\tilde{b}(X, A) | X] = \beta(X),$$

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

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

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

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 latter approach is particularly simple and foreshadows our approach to estimation in the continuous case. When X_t is continuous we impose smoothness restrictions on $\beta(x)$ which are sufficient to point identify 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, the ‘movers’ average partial effect (MAPE) as

$$\beta^M \equiv \mathbb{E}[b(A, U_t) | \Delta X \neq 0] = \frac{\mathbb{E}[\mathbf{1}(\Delta X \neq 0) \beta(X)]}{\mathbb{E}[\mathbf{1}(\Delta X \neq 0)]}. \quad (6)$$

Expression (6) is implicit in Chamberlain (1982, p. 13) who also noted that we have no information on $\beta^S = \mathbb{E}[b(A, U_t) | \Delta X = 0]$, 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).

Consider, to illustrate the main ideas, the case where X_t is binary, for example an indicator for union membership, and $Y_t \in [\underline{y}, \bar{y}]$.¹¹ In that case the APE is given by

$$\beta = \sum_{i,j=0,1} \pi_{ij} \beta(i, j),$$

where $\pi_{ij} = \Pr(X_1 = i, X_2 = j)$. Recall, however, that $\beta(0, 0)$ and $\beta(1, 1)$, the partial effects associated with the two types of stayers are only partially identified with respective identification regions of¹²

$$\mathbb{H}\{\beta(0, 0)\} = [\underline{y} - \mathbb{E}[Y_t | X_1 = 0, X_2 = 0], \bar{y} - \mathbb{E}[Y_t | X_1 = 0, X_2 = 0]]$$

¹¹Note that for X_t binary the CRC structure is unrestrictive (although our marginal stationarity assumption is restrictive). In independent work Chernozhukov, Fernández-Val, Hahn and Newey (2007) develop bounds for β for the case where Y_t is binary. They show that if, in our notation, $\mathbb{E}[Y_t(x_t) | A = a] = \Phi(x_t' \gamma + a)$, tighter bounds are available.

¹²Observe that, from equation (2),

$$\beta(0, 0) = \mathbb{E}[Y_t(1) | X_1 = 0, X_2 = 0] - \mathbb{E}[Y_t(0) | X_1 = 0, X_2 = 0]$$

The data reveal

$$\mathbb{E}[Y_t(0) | X_1 = 0, X_2 = 0] = \mathbb{E}[Y_t | X_1 = 0, X_2 = 0]$$

but do not reveal $\mathbb{E}[Y_t(1) | X_1 = 0, X_2 = 0]$, which may lie anywhere in the interval $[\underline{y}, \bar{y}]$.

and

$$\mathbb{H}\{\beta(1, 1)\} = [\mathbb{E}[Y_t | X_1 = 1, X_2 = 1] - \bar{y}, \mathbb{E}[Y_t | X_1 = 1, X_2 = 1] - \underline{y}].$$

The identified set for the APE, β , is thus

$$\begin{aligned} & \mathbb{H}\{\beta\} \\ & \in [\pi_{00}(\underline{y} - \mathbb{E}[Y_t | X_1 = 0, X_2 = 0]) + \pi_{01}\beta(0, 1) + \pi_{10}\beta(1, 0) + \pi_{11}(\mathbb{E}[Y_t | X_1 = 1, X_2 = 1] - \bar{y}), \\ & \quad \pi_{00}(\bar{y} - \mathbb{E}[Y_t | X_1 = 0, X_2 = 0]) + \pi_{01}\beta(0, 1) + \pi_{10}\beta(1, 0) + \pi_{11}(\mathbb{E}[Y_t | X_1 = 1, X_2 = 1] - \underline{y})]. \end{aligned}$$

The width of $\mathbb{H}\{\beta\}$ is $(\pi_{00} + \pi_{11})(\bar{y} - \underline{y})$: the more ‘stayers’ the less informative the data are for β . Unfortunately a preponderance of stayers is common in many microeconomic applications. In Card’s (1996) analysis of the union wage premium, less than 10 percent of workers switch between collective bargaining coverage and non-coverage across periods (Table V, p. 971). In such cases 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.

Before turning to the continuous case we briefly discuss identification of the LAR when X_t is discrete. As with the APE, the LAR is generally not identified. Instead we can identify the ‘movers’ local average response (MLAR). This is given by

$$\gamma^M(g(x)) \equiv \mathbb{E}[b(A, U_t) | g(X) = g(x), \Delta X \neq 0] = \frac{\mathbb{E}[\mathbf{1}(g(X) = g(x)) \mathbf{1}(\Delta X \neq 0) \beta(X)]}{\mathbb{E}[\mathbf{1}(g(X) = g(x)) \mathbf{1}(\Delta X \neq 0)]}. \quad (7)$$

It is straightforward to construct bounds on $\gamma(g(x))$ along the lines of those given for β above. Note that, by stationarity, $\gamma^M(g(x))$ and $\gamma(g(x))$ do not depend on t .

Continuous regressor When X is continuous the set $\{x : x \in \mathcal{X}, x_1 = x_2\}$ 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

$$\beta(x_1, x_1) = \lim_{h \downarrow 0} \frac{\mathbb{E}[Y_2 | X = (x_1, x_1 + h)] - \mathbb{E}[Y_1 | X = (x_1, x_1)]}{h}. \quad (8)$$

A sufficient condition for the above limit to exist is:

Assumption 2.4 (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, 2.2, 2.3 and 2.4 hold, then $\beta_t(x_t) = \beta$ and $\gamma_t(x_t)$ are identified by*

$$\beta = \mathbb{E}[\beta(X)], \quad \gamma_t(x_t) = \mathbb{E}[\beta(X) | X_t = x_t]$$

with $\beta(x)$ given by (5) or (8) as appropriate.

Proof. Straightforward and therefore omitted. ■

Observe that $\beta(x)$ is an average over the conditional distribution of (A, U_t) given X . Thus smoothness of $\beta(x)$ suggests that the distribution function of A given $X = x$ is smooth in x . Such smoothness conditions are often implied by correlated random effect specifications for A . A fixed effects purist could thus call our model (when X_t is continuous) a correlated random effects one. We maintain the fixed effects characterization because we view Assumption 2.4 as rather weak. In 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 and marginal stationarity imply that:

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

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

$$\{x, x' : x, x' \in \mathcal{X}, \quad x_1 = x_2, \quad x'_1 = x'_2, \quad x_1 \neq x'_1\}.$$

Linearity of $Y_t(x)$ in x , and constancy of the conditional mean of the random slopes over time (given all leads and lags of X_t) means that outcome changes for stayers are driven solely by changes in $a(A, U_t)$ over time. Since marginal stationarity implies constancy of the conditional mean of $a(A, U_t)$, however, 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 generates many testable restrictions.

We can use these extra model restrictions to incorporate aggregate time effects into our model in a fairly flexible way. Consider the model

$$Y_t = a_t(A, U_t) + b(A, U_t) X_t, \tag{9}$$

where $a_t(A, U_t)$ the mapping from A and U_t into the intercept is time-specific but restricted by the following assumption:

Assumption 2.5 (CONDITIONAL COMMON AVERAGE TRENDS)

$$\mathbb{E}[a_2(A, U_t) - a_1(A, U_t) | X] = \delta(X_2).$$

Assumption 2.5 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 2.5 allows for the period two shock to affect mean earnings in the union and non-union sectors differently.

Let $x = (x_1, x_2)'$ with $x_1 \neq x_2$ and $x' = (x'_1, x'_2)'$ with $x'_1 = x'_2 = x_2$. Observe that

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

while

$$\mathbb{E}[Y_2|X = x] - \mathbb{E}[Y_1|X = x] = \delta(x_2) + \beta(x)(x_2 - x_1).$$

We therefore have

$$\beta(x) = \frac{\mathbb{E}[Y_2|X = x] - \mathbb{E}[Y_1|X = x] - \{\mathbb{E}[Y_2|X = x'] - \mathbb{E}[Y_1|X = x']\}}{x_2 - x_1} \quad (10)$$

We may adapt expression (8) to get $\beta(x)$ for stayers. With $\beta(x)$ identified, identification of the APE and LAR follows directly.

Assumption 2.5 is a substantial 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 2.5 is invoked with the additional the requirement that $\delta(X_2) \equiv \delta$ is constant in X_2 ; which is an ‘unconditional’ common average trends assumption.

For estimation purposes it is convenient to assume a parametric form for $\delta(X_2)$. A natural specification, given the CRC form for the outcome response, is $\delta(X_2) = \delta_a + \delta_b X_2$. Note that this specification for $\delta(X_2)$ is isomorphic to the pair of restrictions

$$\mathbb{E}[a_2(A, U_t) - a_1(A, U_t)|X] = \delta_a, \quad \mathbb{E}[b_2(A, U_t) - b_1(A, U_t)|X] = \delta_b,$$

in the model where $b_t(A, U_t)$ the mapping from A and U_t into the slope is also time-specific (i.e., $Y_t = a_t(A, U_t) + b_t(A, U_t)X_t$). Put differently, linearity of $\delta(X_2)$ is equivalent to a CRC model which allows for aggregate common intercept and slope drift across periods. 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

$$\frac{\mathbb{E}[\Delta Y|X = x''] - \mathbb{E}[\Delta Y|X = x']}{x''_2 - x'_2} = \delta_b.$$

We work with this specification for $\delta(X_2)$ in the next subsection and with the even simpler constant in X_2 aggregate effect $\delta(X_2) = \delta$ 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) estimator

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

$$Y_t = \delta_t + \beta X_t + A + U_t, \quad \mathbb{E}[U_t | A, X] = 0, \quad t = 1, 2 \quad (11)$$

when in fact the true model is as described by Assumptions 2.1, 2.2, 2.3, 2.4 and 2.5 with $\delta(X_2) = \delta_a + \delta_b X_2$.

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

$$b^{FE} = \beta + \delta_b \left\{ 1 - \frac{\mathbb{C}(X_1, X_2)}{\mathbb{V}(\Delta X)} \right\} + \mathbb{E}[\omega(\Delta X)(\beta(X) - \beta)], \quad \omega(\Delta X) = \frac{\Delta X (\Delta X - \mathbb{E}[\Delta X])}{\mathbb{V}(\Delta X)}. \quad (12)$$

The first term in (12) 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 (12) and the LATE is that ‘movers’, unlike ‘compliers’, can be directly identified from the data. Consequently the weights in (12) 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_1^2 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

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

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 (11) does not include time-specific intercepts).¹³

The linear FE estimator is especially interpretable in the ‘classical’ difference-in-differences

¹³In independent work, Chernozhukov, Fernández-Val, Hahn and Newey (2007, Section 3.3) obtain a related result in the context of a fixed effects binary choice model (without time effects).

(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).

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).¹⁴ 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 (12) 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 with 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.

2.3 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

$$Y_t = m_t(X_t, A, U_t) \tag{13}$$

and the following exchangeability assumption:

Assumption 2.6 (EXCHANGEABILITY) (i)

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

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

¹⁴Wooldridge (2005a) also assumes that the correlated random coefficients are time invariant.

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

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

where $\zeta_t(X)$ is the t^{th} elementary symmetric polynomial on X .¹⁵ Because Assumption 2.6 is not sufficient to identify $\beta_t(x)$ Altonji and Matzkin (2005, pp. 1063 - 4) suggest either further restricting the conditional distribution of (A, U_t) given X or the form of the structural outcome equation.¹⁶

Following their second suggestion, the imposition of our CRC structure on (13) and Assumption 2.6 implies that

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

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 2.6 nest the other. For example, while Assumption 2.2 does not restrict the conditional distribution of A given X it does exclude time-varying heteroscedasticity allowed by Assumption 2.6.

A natural combination of the two assumptions is:

Assumption 2.7 (STATIONARITY AND EXCHANGEABILITY) (i)

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

¹⁵These polynomials take the form $\zeta_1(X) = \sum_{1 \leq i \leq T} X_i$, $\zeta_2(X) = \sum_{1 \leq i < j \leq T} X_i X_j$, $\zeta_3(X) = \sum_{1 \leq i < j < k \leq T} X_i X_j X_k$, $\zeta_4(X) = \sum_{1 \leq i < j < k < l \leq T} X_i X_j X_k X_l$ and so on up to $\zeta_T(X) = \prod_{i=1}^T X_i$.

¹⁶One 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).

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

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

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

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

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

when x' is a permutation of x .

3 Estimation

In this section we discuss estimation of the movers average partial effect, β^M , and movers local average response, $\gamma_t^M(x_t)$, when the regressors are discretely-valued, and the average partial effect, β , and local average response, $\gamma_t(x_t)$, with continuously-valued regressors. To keep the exposition simple we work with the constant in X_2 aggregate time effects specification $\delta(X_2) = \delta$.

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

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

We, of course, require that $\Pr(\Delta X = 0)$ is greater than zero: *it is the presence of stayers which identifies δ_0* . Now consider the subpopulation of movers, we have

$$\mathbb{E}[\Delta Y | X = x] = \delta_0 + \beta(x) \Delta x,$$

and hence, with δ identified, we may write

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

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

$$\psi(Z, \theta) = \begin{pmatrix} \mathbf{1}(\Delta X = 0) (\Delta Y - \delta) \\ \frac{\mathbf{1}(\Delta X \neq 0)}{\Delta X} (\Delta Y - \delta - \beta^M \Delta X) \end{pmatrix}.$$

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 $\mathbf{1}(\Delta X = 0)$ and $\frac{\mathbf{1}(\Delta X \neq 0)}{\Delta X}$ serving as excluded instruments (this follows since $\mathbf{1}(\Delta X = 0)(\Delta Y - \delta - \beta^M \Delta X) = \mathbf{1}(\Delta X = 0)(\Delta Y - \delta)$). 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 $\Gamma_0 = \mathbb{E}[\partial\psi(Z, \theta_0)/\partial\theta']$ and $\Omega_0 = \mathbb{E}[\psi(Z, \theta_0)\psi(Z, \theta_0)']$ and further define

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

We have

$$\Gamma_0 = -\begin{pmatrix} \pi_0 & 0 \\ (1 - \pi_0)\xi & (1 - \pi_0)\kappa \end{pmatrix}, \quad \Omega_0 = \begin{pmatrix} \pi_0\sigma_0^2 & 0 \\ 0 & (1 - \pi_0)\kappa \end{pmatrix},$$

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

$$\sqrt{N}\begin{pmatrix} \widehat{\delta} - \delta \\ \widehat{\beta}^M - \beta^M \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \frac{\sigma_0^2}{\pi_0} & -\frac{\sigma_0^2}{\pi_0}\xi \\ -\frac{\sigma_0^2}{\pi_0}\xi & \kappa + \frac{\sigma_0^2}{\pi_0}\xi^2 \end{pmatrix}\right). \quad (14)$$

Two features of (14) 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. Recall, however, that as π_0 increases so does the width of our bounds on β . This is a manifestation of the tension between the identifying power of stayers for common trends, and the problems they create for identification of the APE.

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 . If instead $\xi = 0$, the asymptotic properties of $\widehat{\beta}^M$ do not depend on those of $\widehat{\delta}$. In that case we can estimate $\widehat{\beta}^M$ as precisely as we could if we somehow knew δ *a priori*.

Estimation of the MLAR is also straightforward. An argument analogous to that given for β^M yields the representation

$$\gamma^M(g(x)) = \frac{\mathbb{E}\left[\mathbf{1}(g(X) = g(x))\mathbf{1}(\Delta X \neq 0)\frac{\Delta Y - \delta}{\Delta X}\right]}{\mathbb{E}\left[\mathbf{1}(g(X) = g(x))\mathbf{1}(\Delta X \neq 0)\right]}.$$

We may therefore estimate $\theta = (\delta, \gamma^M(g(x)))'$ by the analog estimator based on the population

moment restriction $\mathbb{E}[\psi(Z, \theta_0)] = 0$ with

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

We, of course, require that $\Pr(g(X) = g(x), \Delta X \neq 0) > 0$.

3.2 Continuous regressor case

When X_t is continuously valued then, under smoothness conditions, the APE and LAR are 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 and $\gamma(g(x))$ generally converge at the one- and two-dimensional nonparametric rates respectively (assuming $g(x)$ is scalar-valued). 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. We then discuss estimation of $\gamma(g(x))$.

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

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

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

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

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 $\tilde{\beta}$ 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

$$\hat{\beta}(h_N) \equiv \frac{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N) \left(\frac{\Delta Y_i}{\Delta X_i} \right)}{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)}, \quad (15)$$

where h_N is a deterministic bandwidth sequence tending to zero as N tends to infinity.¹⁷

The estimator $\widehat{\beta}(h_N)$ – which is consistent for β^M when X has finite support – has asymptotic properties similar to a standard (uniform) kernel regression estimator for a one-dimensional problem. In particular, it is straightforward to verify that

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

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

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

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

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

¹⁷An alternative consistent estimator would replace the denominator by the sample size N .

Assuming $p = 2$, the asymptotic distribution of $\widehat{\beta}$ is similar as 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 with the density ϕ_0 .

3.2.2 Aggregate time effect

When aggregate time effects are present, and the ‘common trends’ condition (Assumption 2.5) holds with $\delta(X_2) = \delta$, then (10) implies that the average partial effect β is identified by

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

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

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

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

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

$$\widehat{\delta} \equiv \frac{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| \leq h_N) \Delta Y_i}{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| \leq h_N)}, \quad (16)$$

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

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

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

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

$$\widehat{\beta}_F = \frac{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N) \left(\frac{\Delta Y_i - \widehat{\delta}}{\Delta X_i} \right)}{\sum_{i=1}^N \mathbf{1}(|\Delta X_i| > h_N)}. \quad (17)$$

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

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

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

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

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 $\widehat{\xi}$ 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 the estimator $\widehat{\delta}$ of the common trend. An intermediate case could have $\tau(d) = O(d)$ in a neighborhood of zero, with the third term converging in probability to some nonzero limit.

In any event, an asymptotic variance estimator for $\widehat{\beta}_F$ 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 $\widehat{\delta}$ can be constructed. Under standard regularity conditions, the kernel estimators

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

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, and will be continuously distributed elsewhere. If there is overlap in the mass points of X_1 and X_2 , then the distribution of first differences ΔX will generally have a mass point at zero, and will otherwise be continuously distributed in a neighborhood of zero. In this setting, the

average partial effect 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 (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 (15) and (17) above, when ‘stayers’ are a non-negligible portion of the population.

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

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

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

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

so that preliminary estimation of the common trend component δ will not affect the asymptotic distribution of the feasible estimator $\hat{\beta}_F$. 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

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

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

Defining

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

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.,

$$\begin{pmatrix} \bar{\delta} \\ \hat{\beta}^S \end{pmatrix} = \arg \min_{d, b^S} \sum_{i=1}^N \mathbf{1}(|\Delta X_i| \leq h_N) \cdot (\Delta Y_i - d - b^S \Delta X_i)^2, \quad (18)$$

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

3.2.4 Local average response

Let $\mathcal{K}(\cdot)$ be some symmetric function which integrates to one and possibly satisfies other conditions. A consistent estimate of the local average response, $\gamma(g(x))$, is

$$\hat{\gamma}(g(x)) = \frac{\sum_{i=1}^N \mathcal{K}\left(\frac{g(X_i) - g(x)}{h_{2N}}\right) \mathbf{1}(|\Delta X_i| > h_{1N}) \left(\frac{\Delta Y_i - \hat{\delta}}{\Delta X_i}\right)}{\sum_{i=1}^N \mathcal{K}\left(\frac{g(X_i) - g(x)}{h_{2N}}\right) \mathbf{1}(|\Delta X_i| > h_{1N})}$$

where $\hat{\delta}$ is given by (16) above.

3.2.5 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 those defined by (18) above, while the coefficient on the last regressors is equal to (17) (with $\bar{\delta}$ replacing $\hat{\delta}$). The robust standard errors reported by most statistical packages will be asymptotically valid.¹⁸

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 $\hat{\beta} \equiv \hat{\pi}\hat{\beta}^S + (1 - \hat{\pi})\hat{\beta}_F$. A combination of the conventional GMM covariance matrix and the textbook delta method may be used to form standard errors for $\hat{\beta}$.

4 Multiple regressors and time periods

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

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

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

¹⁸Note 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).

Our marginal stationarity restriction is

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

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

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

and

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

To complete the model we make the additional restrictions that

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

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 $\mathbf{d}(A, U_t)$ 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 with $T = p = 2$ and $q = 1$. The matrix of time shifters and its corresponding coefficient vector parameterize the common intercept shift across periods:

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

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 $\boldsymbol{\delta} \equiv \mathbb{E}[\mathbf{d}(A, U_t)]$, $\boldsymbol{\beta} \equiv \mathbb{E}[\mathbf{b}(A, U_t)]$ and $\boldsymbol{\gamma}_t(g(\mathbf{x})) \equiv \mathbb{E}[\mathbf{b}(A, U_t) | g(\mathbf{X}) = g(\mathbf{x})]$.

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 $\boldsymbol{\beta}$ 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 $\mathbf{Y} = (Y_1, \dots, Y_T)'$ we have

$$\mathbb{E}[\mathbf{Y}|\mathbf{W}, \mathbf{X}] = \mathbf{W}\boldsymbol{\delta} + \mathbf{X}\boldsymbol{\beta}(\mathbf{X}). \quad (19)$$

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} ,

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

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 $i = 1, \dots, N$, 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

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

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 $\boldsymbol{\delta}$; 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

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

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

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

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 $\tilde{X} = \det(\mathbf{X}) \neq 0$.

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 $\hat{\boldsymbol{\beta}}_F$ to this context is

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

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

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

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

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

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

In the mixed discrete-continuous case $\Pr(\tilde{X} = 0) > 0$, and estimation of $\hat{\boldsymbol{\beta}}$ 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 typically 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 (19) 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

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

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

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

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

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

which implies that α and δ 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 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

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

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 $\mathbf{W} = \boldsymbol{\omega}(\mathbf{X})$, 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

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

for $\hat{\boldsymbol{\delta}}$ 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 $\hat{\boldsymbol{\beta}}_i$ 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

$$\mathbb{E} \left[\frac{1}{\det(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})} \right] < \infty. \quad (20)$$

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 (20) 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 $X_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ and $\mathbf{V}_i \equiv \mathbf{I}$; then

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

and (20) 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 (20) 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 (20) 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 $\hat{\boldsymbol{\beta}}_i$ 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 (20) holds, and $h_N = o(1/\sqrt{N})$. Since $\hat{\boldsymbol{\beta}}$ will still be consistent for β even when (20) fails, a feasible version of the trimmed mean $\hat{\boldsymbol{\beta}}$ 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 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 the 2001 and 2002 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 INEC (2005) we converted all food quantities into grams. We then used the caloric content and edible percent information in the INCAP (2000) food composition tables to construct a measure of daily total 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

$$I_{it} = \left[S_{it} \left\{ \sum_{f=1}^F W_{f,it} \left(P_f^b / P_{f,it} \right) \right\} + (1 - S_{it}) J_{it} \right]^{-1}, \quad (21)$$

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). To avoid the introduction

¹⁹A 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 are available online at <http://www.econ.berkeley.edu/~bgraham/>.

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

of spurious noise we, following the suggestion of Deaton and Zaidi (2002), replace household-level unit prices with village medians. 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 2002. Our independent variable of interest is the logarithm of real per capita consumption: $X_{it} = \ln([C_{it}/I_{it}]/M_{it})$; M_{it} is total household size.

Table 1 summarizes some key features of our estimation sample. Panel A gives the share of total food spending devoted to each of eleven broad food categories as well as the share of total calories derived from each group. 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. 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

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 2 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).

²¹In 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 (%)						Calorie Shares (%)					
	All		Lower 25%		Upper 25%		All		Lower 25%		Upper 25%	
	2001	2002	2001	2002	2001	2002	2001	2002	2001	2002	2001	2002
Cereals	36.0	32.7	41.0	35.6	32.2	29.4	60.3	59.9	63.7	61.9	58.1	58.2
Roots	3.1	2.7	2.3	1.7	3.5	3.8	1.5	1.6	1.4	1.2	1.7	2.3
Pulses	12.5	13.6	14.6	16.6	10.1	10.9	11.3	12.8	11.8	13.6	10.5	11.7
Vegetables	4.9	4.5	4.0	3.2	5.8	5.6	0.7	0.6	0.5	0.4	0.9	0.8
Fruit	0.9	1.1	0.6	0.9	1.1	1.3	0.5	0.4	0.2	0.3	0.7	5.8
Meat	6.9	7.7	3.6	4.7	9.7	10.6	1.3	1.3	0.6	0.7	2.1	1.9
Dairy	14.7	17.3	11.5	15.5	16.2	19.5	4.3	3.7	1.7	2.2	4.6	5.0
Oil	5.0	5.0	5.4	5.2	4.0	4.5	7.6	7.1	6.2	5.9	7.6	7.6
Other foods	16.0	15.4	17.0	16.5	14.2	14.5	12.6	10.3	9.5	9.8	11.8	10.5
Staples [◇]	51.6	49.0	57.8	54.0	45.7	44.0	73.1	74.3	76.9	76.8	70.3	72.1
Panel B:	Total Real Expenditure & Calories											
Expenditure per adult ^b	4,679	4,510	2,225	2,061	7,766	7,773						
(Expenditure per capita)	(3,764)	(3,887)	(1,947)	(1,991)	(5,970)	(6,320)						
Food share	69.2	68.8	68.4	68.5	68.7	68.2						
Calories per adult ^b	3,015	2,949	2,045	1,952	3,981	3,878						
(Calories per capita)	(2,435)	(2,530)	(1,776)	(1,825)	(3,050)	(3,121)						
Percent energy deficient [‡]	39.3	39.7	73.5	77.6	12.7	10.6						

Table 1: Real food expenditure and calorie shares of RPS households in 2001 and 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 nominal 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). The Data Appendix provides complete details. Lower and upper 25 percent refers to the bottom and top quartiles of households based on the sum of 2001 and 2002 real consumption per adult equivalent and thus contains the same set of households in both years.

[◇] Sum of cereal, roots and pulses.

^b "Adults" correspond to adult equivalents based on FAO (2001) recommended energy requirements for light activity.

[‡] Percentage of households with estimated calorie availability less than FAO (2001) recommendations for light activity given household demographics.

	Median Cordobas per 1,000 calories					
	All		Bottom 25%		Top 25%	
	2001	2002	2001	2002	2001	2002
Cereals	1.3	1.2	1.2	1.0	1.5	1.4
Roots	8.6	7.2	7.6	6.0	8.6	7.2
Pulses	2.6	2.3	2.6	2.4	2.6	2.6
Vegetables	23.2	22.7	22.7	19.6	24.2	23.8
Fruit	6.3	6.6	5.4	5.5	6.6	7.1
Meat	19.1	18.6	18.6	18.1	19.9	19.3
Dairy	10.1	10.0	10.9	10.4	10.3	10.1
Oil	1.5	1.5	1.6	1.5	1.5	1.5
Other foods	3.1	3.1	2.9	2.8	3.4	3.6
All foods	2.4	2.4	2.0	2.0	3.0	2.9

Table 2: Real Cordobas spent by RPS households in 2001 and 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 sum of 2001 and 2002 real consumption per adult equivalent. See notes to Table 1 and the Data Appendix for additional details.

5.2 Model and estimation

We assume that the logarithm of total calorie availability per capita in period t , Y_t , varies according to

$$Y_t = W_{1t}'\delta_1 + a_t(A, U_t) + b_t(A, U_t)X_t, \quad (22)$$

where W_{1t} is a vector of time-varying demographic controls capturing the age and gender composition of the household, X_t is the logarithm of real total outlay per capita, and $a_t(A, U_t)$ and $b_t(A, U_t)$ are a household- and time-specific intercept and calorie elasticity. We allow for common intercept and elasticity drift of the form:

$$\mathbb{E}[a_2(A, U_2) - a_1(A, U_1) | W_1, X] = \delta_a, \quad \mathbb{E}[b_2(A, U_2) - b_1(A, U_1) | W_1, X] = \delta_b.$$

Defining $\mathbf{W}_1 = (W_{11}, W_{12})'$ with

$$\mathbf{W}_2 = \begin{pmatrix} 0 & 0 \\ 1 & X_2 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \end{pmatrix},$$

and $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2)$ 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 $\boldsymbol{\delta} = (\delta_1', \delta_a, \delta_b)'$ and $\boldsymbol{\beta}(\mathbf{x}) = (\alpha(x), \beta(x))' = (\mathbb{E}[a_1(A, U_t) | X = x], \mathbb{E}[b_1(A, U_t) | X = x])'$.

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. Importantly, it also allows these household-specific elasticities to co-vary with total outlay / income. By virtue of linearity, however, a household’s elasticity does not (structurally) vary with income. Subramanian and Deaton (1996), in contrast, do allow a household’s calorie elasticity to (structurally) vary with income. However this nonlinear mapping is homogenous across all households. This difference merits emphasis: in our model, from the perspective of the household, the elasticity of calorie demand *does not* vary with income. If the household’s elasticity is 0.3, then a ten-fold exogenous increase in income results in a three-fold increase in caloric intake. Clearly structural linearity is a biological impossibility and (22) must necessarily be viewed as an approximation.

To make the contrast between the two modelling approaches explicit, assume that the true model is given by (22) with, for simplicity, $\delta_1 = 0$. Subramanian and Deaton (1996), in the absence of panel data, use a single cross-section to estimate the calorie elasticity.²² Their estimate is the sample analog of the derivative of the conditional expectation function $\mathbb{E}[Y_t | X_t = x_t]$, which under (22) equals

$$\frac{\partial \mathbb{E}[Y_t | X_t = x_t]}{\partial x_t} = \beta(X_t) + X_t \beta'(X_t),$$

where, in an abuse of our established notation, $\beta(X_t) = \mathbb{E}[b_t(A, U_t) | X_t]$. The first term is structural, while the second reflects heterogeneity bias. If $\beta(X_t)$ does indeed decline with income as is commonly conjectured (i.e., $\beta'(X_t) < 0$), then their estimate of the calorie elasticity will be biased downwards in large samples (since $X_t > 0$). The procedure will also exaggerate the degree to which the elasticity changes with income.

Our point is not to argue for the superiority of one approach or the other, but to highlight the substantive differences between them. At the cost of linearity, 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.²³

For computation we employ an ‘instrumental variables’ procedure. Our estimates of π , δ , β^S and β^M are given by the solution to (suppressing the dependence of our estimator on the choice of bandwidth)

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

²²We 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.

²³For example, with three years of panel data, we could allow each household’s elasticity to decline (linearly) with log-income.

with $\hat{\theta} = (\hat{\pi}, \hat{\delta}', \hat{\beta}^S, \hat{\beta}^M)'$ and

$$\psi_i(\theta) = \left(\begin{array}{c} \mathbf{1}(|\tilde{X}_i| \leq h_N) - \pi \\ \mathbf{z}_i' \left(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{W}}_i \delta - \tilde{X}_i \left(\mathbf{1}(|\tilde{X}_i| \leq h_N) \cdot \mathbf{I}_p \quad \mathbf{1}(|\tilde{X}_i| > h_N) \cdot \mathbf{I}_p \right) \left(\begin{array}{c} \beta^S \\ \beta^M \end{array} \right) \right) \end{array} \right),$$

where

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

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

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

The distribution of X_t does not appear to have discrete components, therefore $\hat{\beta}^M$ should provide a consistent estimate of the overall average elasticity of calorie demand. Nevertheless we also report (and prefer) estimates of β based on

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

We compute standard errors using the conventional GMM variance-covariance estimator (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.

We use several different values of the bandwidth h_N ; our default choice is the ‘rule-of-thumb’ choice $h_N = \hat{\sigma}_{\tilde{X}} \cdot N^{-1/5}/2$, with $\hat{\sigma}$ the estimated standard deviation of $\tilde{X}_i \equiv \det(\mathbf{X}_i)$ ($= 0.5315$). This yields a bandwidth value of 0.0659. We also consider the bandwidth that minimizes the ‘cross-validation’ criterion

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

where $\hat{\theta}_{-i}$ is the estimate calculated by omitting the i^{th} observation from the sample.

5.3 Results

Table 3 reports pooled OLS and linear fixed effects estimates of the calorie elasticity. Controls for the age and sex composition of the household are also included (coefficients not reported)²⁴. 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 calorie elasticity, to shift across

²⁴The demographic controls are the total number of household members in each of sixteen age-by-sex categories.

	Pooled OLS			FE-OLS		
	(1.a)	(1.b)	(1.c)	(2.a)	(2.b)	(2.c)
\widehat{b}	0.5592 (0.0353)	0.5592 (0.0356)	0.5761 (0.0496)	0.6357 (0.0392)	0.6351 (0.0390)	0.6613 (0.0480)
	Aggregate Time Effects			Aggregate Time Effects		
\widehat{d}_a	—	0.0274 (0.0273)	0.2870 (0.3592)	—	0.0308 (0.0287)	0.4227 (0.3657)
\widehat{d}_b	—	—	−0.0321 (0.0426)	—	—	−0.0485 (0.0439)
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Table 3: Conventional 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 2001 and 2002 samples, "FE-OLS" denotes least squares estimates with household-specific intercepts. Included demographic controls are the number of individuals in a household in each of sixteen age-by-gender categories. The standard errors are computed in a way that allows for arbitrary within-village correlation in disturbances across households and time.

periods. Relative to the pooled OLS results, those which include household-specific intercepts are larger in magnitude by about 15 percent (e.g., 0.56 versus 0.64 in Columns 1.a versus 2.a). 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 4 reports estimates based on the CRC model. Column 1 is the simple Mundlak/Chamberlain estimator $\widehat{\beta} = N^{-1} \sum_{i=1}^N \Delta Y_i / \Delta X_i$. Consistency of this estimator requires that $\mathbb{E}[|\Delta Y / \Delta X|] < \infty$ (Chamberlain, 1980b, 1982). Given that the density of ΔX is substantial in the neighborhood of zero (see Figure 1), this condition is unlikely to be satisfied in the present setting. This estimator also precludes the presence of aggregate time effects and additional demographic controls. In any case it produces a small, and very imprecisely estimated, calorie elasticity.

Columns 2 through 4 report results based on our estimator using the same ‘rule-of-thumb’ bandwidth in each column ($h_N = 0.0659$). Column 2 excludes aggregate times effects, while Columns 3 and 4 allow for, respectively, intercept, and intercept and elasticity, drift over time. The average calorie elasticity estimated by our procedure is approximately 0.4 in each case. This is less than two-thirds the value implied by the linear fixed-effects estimator (although a 95 percent confidence interval easily contains the FE point estimate).

To assess the importance of the choice of bandwidth for our results Table 5 re-estimates the Table 4, Column 2 specification using several alternative bandwidth choices. If we decrease the bandwidth to three-quarters of the rule-of-thumb value the CRC estimates are nonsensical and imprecise. This reflects a combination of ‘zero denominator’ problems and the difficulty of precisely estimating the sixteen coefficients on the household demographic variables using only 5 percent of the sample. Column 3 and 4 reports estimates based on 1.5 times the rule-of-thumb bandwidth; here the point estimate, while statistically indistinguishable from that reported in Column 1, is larger in

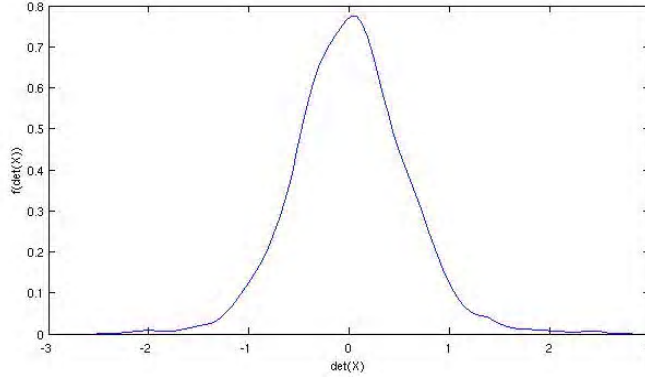


Figure 1: Kernel density estimate of the density of $\det(\mathbf{X}) = \Delta X$.

NOTES: Density estimate based on a Gaussian kernel and Silverman’s rule-of-thumb bandwidth.

	CRC-Panel			
	(1)	(2)	(3)	(4)
β^S	—	0.4503 (1.13036)	0.5339 (1.1469)	0.5431 (1.1430)
β^M	—	0.3929 (0.2987)	0.3914 (0.3006)	0.4354 (0.2987)
β	0.1746 (1.1418)	0.3990 (0.1324)	0.4067 (0.1587)	0.4470 (0.1492)
	Aggregate Time Effects			
δ_a	—	—	-0.0174 (0.0476)	0.3901 (0.6181)
δ_b	—	—	—	-0.0500 (0.0735)
Demographics	No	Yes	Yes	Yes

Table 4: 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 same ‘rule-of-thumb’ bandwidth value is used across columns 2 to 4, $h_n = 0.5 \cdot \sigma_{\Delta X} \cdot N^{-1/5} = 0.065925$. This corresponds to ‘trimming’ 10.7 percent of the sample. Demographic controls include the number of household members in each of sixteen age-by-gender categories. The standard errors are computed in a way that allows for arbitrary within-village correlation in disturbances across households and time.

	CRC-Panel			
	(1)	(2)	(3)	(4)
β^S	0.4503 (1.13036)	-1.8756 (1.6002)	0.2647 (0.4853)	0.5648 (0.0522)
β^M	0.3929 (0.2987)	4.1578 (2.6494)	0.5940 (0.1143)	0.6334 (0.0400)
β	0.3990 (0.1324)	3.6913 (3.0564)	0.5438 (0.2245)	0.5887 (0.0741)
Demographics	Yes	Yes	Yes	Yes
h_N	0.0659	0.0494	0.0989	0.48
Percent 'trimmed'	10.7	7.7	15.2	65.2

Table 5: Correlated random coefficient estimates of the elasticity of calorie demand with respect to household expenditure: the effects of bandwidth choice

NOTES: Estimates based on the balanced panel of 1,358 households described in the main text. Column 1 gives the Column 2 estimates listed in the previous table. Columns 2 and 3 report estimates of the same specification with bandwidths equal to, respectively, 3/4 and 1.5 times the value of the Column 1 ‘rule-of-thumb’ bandwidth. Column 4 reports estimates based on the bandwidth selected by the cross-validation procedure described in the main text. Demographic controls include the number of household members in each of sixteen age-by-gender categories. The standard errors are computed in a way that allows for arbitrary within-village correlation in disturbances across households and time.

magnitude. Finally Column 4 reports the estimate based on the cross-validation-chosen bandwidth (Figure 2 plots the leave-own-out quadratic form criterion as a function of the bandwidth for this model). Cross-validation suggests the use of a significantly larger bandwidth.²⁵ In this case the estimated average elasticity is less than ten percent smaller than its linear fixed effects counterpart (0.59 versus 0.64).

Overall our results are consistent with a modest positive average elasticity of calorie demand. Relative to OLS estimates which include household-specific intercepts, our CRC estimates are systematically smaller in magnitude. These differences are not, however, statistically significant. The exact magnitude of the difference is sensitive to the choice of bandwidth.

5.4 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 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. Our estimates are suggestive, albeit not decisively so, of the presence of correlated heterogeneity. They are consistently smaller in absolute value than conventional linear fixed effects estimates. 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 relatively little guidance on

²⁵ Although Figure 2 indicates that the criterion is relatively flat over a large range of bandwidth choices.

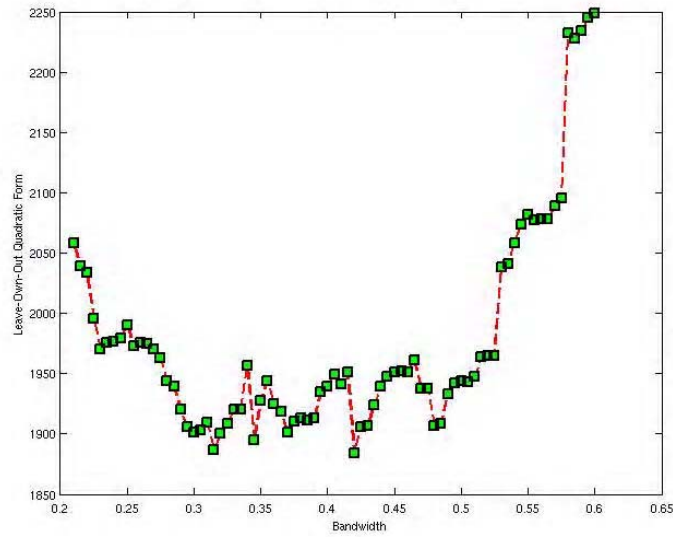


Figure 2: Leave-own-out ‘IV’ quadratic form for different bandwidth values – Column 7, Table 5 specification.

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 $b_t(A, U_t)$), to accommodate additional ‘triangular endogeneity’ and to allow for discrete and/or censored outcome variables.

Appendices

A Derivations

[TO BE DRAFTED]

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," *Investigaciones 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 revisited," *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," *NBER Working Paper 14002*.
- 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 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 Progreso 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.
- 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. (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.