Estimation of partial effects in non-linear panel data models*

by Jason Abrevaya† and Yu-Chin Hsu‡

This version: September 2011

ABSTRACT

Nonlinearity and heterogeneity complicate the estimation and interpretation of partial effects. This paper provides a systematic characterization of the various partial effects in non-linear panel-data models that might be of interest to empirical researchers. The estimation and interpretation of the partial effects depends upon (i) whether the distribution of unobserved heterogeneity is treated as fixed or allowed to vary with covariates and (ii) whether one is interested in particular covariate values or an average over such values. The characterization covers partial-effects concepts already in the literature but also includes new concepts for partial effects. A simple panel-probit design highlights that the different partial effects can be quantitatively very different. An empirical application to panel data on health satisfaction is used to illustrate the partial-effects concepts and proposed estimation methods.

Keywords: Non-linear panel data models; partial effects; correlated random effects.

---

*We thank participants of the 20th Annual Midwest Econometrics Group for their useful comments. The first author acknowledges financial support from the National Science Foundation (grant SES-0921208). Sample *Stata* code for the methods described in this paper are available on-line at `http://yuchinhsu.yolasite.com/research.php`.

†Address: The University of Texas at Austin, Department of Economics, Austin, TX 78712; e-mail: `abrevaya@eco.utexas.edu`.

‡Address: University of Missouri at Columbia, Department of Economics, Columbia, MO 65211; e-mail: `hsuy@missouri.edu`.

1 Introduction

Estimation of partial effects is important for empirical research. By "partial effect," we mean the effect of a covariate on the outcome holding other covariates fixed. These partial effects can depend on the values of the other covariates, as in a linear regression model with interactions or in nonlinear regression models (e.g. probit). In such cases, researchers often report either an average partial effect or a partial effect evaluated at certain values of the other covariates.

The presence of unobserved heterogeneity in nonlinear panel-data models complicates matters further.[1] In these models, the partial effects are functions of not only the covariates but also the heterogeneity. This dependence leads to an inherent choice about how the heterogeneity should be treated in partial-effects estimation. Specifically, if one wants the effect of a covariate holding other covariates *and* heterogeneity fixed, what does it mean to "hold heterogeneity fixed"?

One approach is to consider the effect of making the same covariate shift for the entire population, which amounts to averaging the partial effect over the unconditional distribution of the heterogeneity. This idea first appeared in Chamberlain (1984) in the context of panel probit models and is related to the *average structural function (ASF)* approach of Blundell and Powell (2003). These partial effects are recommended in the Wooldridge (2010) textbook. See also Wooldridge (2005b) for a more detailed discussion.

An alternative approach is to account for possible correlation between observed covariates and heterogeneity when partial effects are estimated. Specifically, for a covariate value of interest, estimation would consider the *conditional* distribution of heterogeneity given that covariate value. This approach has been proposed in a nonparametric context by Altonji and Matzkin (2005), with the resulting partial effect called a *local average response (LAR)* by the authors.

This paper provides a systematic treatment of these two approaches, focusing on four different notions of partial effects that are summarized in Table 1. The first column, where heterogeneity is treated as a fixed distribution, corresponds to ASF-type partial effects, and the second column, where the conditional distribution of heterogeneity is used, corresponds to LAR-type partial effects. Either of these effects can be evaluated at specific covariate values (corresponding to the "conditional" effects in the first row)

[1] In a *linear* panel-data model, the heterogeneity is generally additively separable and, therefore, does not appear in the partial effect.

Table 1: Terminology for different partial effects

	Distribution of unobserved heterogeneity: treated as a fixed distribution	Distribution of unobserved heterogeneity: allowed to vary with covariates
Estimate evaluated at specific covariate values	*conditional average partial effect* (CAPE)	*conditional average local response* (CALR)
Estimates averaged over covariate values	*average partial effect* (APE)	*average local response* (ALR)

or averaged over covariate values (corresponding to the second row). As will be seen in the discussion below, the *conditional average partial effect (CAPE)* is the effect originally considered by Chamberlain (1984), and the *average partial effect (APE)* is the population average of this effect. The *conditional average local response (CALR)* is a version of LAR that conditions upon a covariate component, whereas the *average local response (ALR)* is the population average of LAR.

Section 2 provides details on the partial effects summarized in Table 1. For expositional purposes, everything is considered within the context of a binary-choice panel-data, but the results and ideas extend naturally to other non-linear models. For the newly introduced LAR-related partial effects, asymptotic results are provided. The focus of this paper is static non-linear panel-data models (models without lagged outcomes as explanatory variables). Carro (2007) discusses different partial effects for the dynamic case, although the estimators in that paper require a large number of time periods ($T \rightarrow \infty$ asymptotics). Section 3 considers a simple data-generating process for the binary-choice model in order to highlight the substantial quantitative differences among the different partial-effect quantities. The section also provides Monte Carlo results that illustrate the finite-sample performance of the various partial-effects estimators. Section 4 compares the various partial-effects notions in the context of an empirical application to panel data on health satisfaction.

2 Partial effects in non-linear panel-data models

The partial-effects concepts and estimation methods are illustrated throughout this paper in the context of a binary-choice panel-data model. We stress, however, that all of the concepts and methods generalize to other non-linear panel-data models. Suppose that for each cross-sectional unit, data are observed for T time periods (where T is a small, fixed number). Consider the classical threshold-crossing binary-choice model with binary outcomes y_t (for $t = 1, \ldots, T$) satisfying

$$y_t = 1(x_t\beta + c + u_t > 0), \tag{1}$$

where x_t is a K-dimensional row vector, c is unobservable heterogeneity that may be related to the covariates $x \equiv (x_1, \ldots, x_T)'$, and u_t is an error disturbance independent of both x and c. Let $u \equiv (u_1, \ldots, u_T)'$ and $y \equiv (y_1, \ldots, y_T)'$ be $T \times 1$ vectors of disturbances and outcomes, respectively.

To fix ideas, consider estimation of partial effects at time period t for a component x_{tj} of the vector x_t.[2] The conditional expectation of the outcome, given time-t covariates and heterogeneity, is denoted

$$m^t(x_t, c) = E_{u_t}[y_t|x_t, c] \tag{2}$$

and the partial effect with respect to x_{tj}, evaluated at given values of the covariates and the heterogeneity, is denoted

$$\theta_j^t(x_t, c) = \partial m^t(x_t, c)/\partial x_{tj}. \tag{3}$$

Let $M_j^t(x_t, x)$ denote the expectation of $\theta_j^t(x_t, c)$ with respect to the conditional distribution of c on x (denoted by $H_c(c|x)$):

$$M_j^t(x_t, x) = E_{c|x}[\theta_j^t(x_t, c)|x] = \int \theta_j^t(x_t, c)dH_c(c|x),$$

which evaluates the average partial effect of changing x_{tj} but conditions the distribution of c on x.

[2] We implicitly assume that x_{tj} enters in a single term of the linear index $x_t\beta$. Dealing with additional terms, such as interaction variables, that include x_{tj} would simply involve application of the chain rule.

2.1 Average local response (ALR)

Note that $M_j^t(x_t, x)$ is the local average response (LAR) parameter introduced by Altonji and Matzkin (2005). Unfortunately, this LAR parameter is difficult to calculate since it depends on x; specifically, in the nonparametric context of Altonji and Matzkin (2005), estimation will be subject to the curse of dimensionality. We instead propose a partial effect that involves an average of the LAR effect (over the distribution of x_t). The *average local response (ALR)*, with respect to x_{jt}, is defined as

$$ALR_j^t = E_x[M_j^t(x_t, x)] = \int M_j^t(x_t, x) dG_x(x) \tag{4}$$

where $G_x(x)$ is the cdf of x. In conjunction with additional parametric assumptions made below, the ALR avoids the curse of dimensionality and is estimable at the parametric $\sqrt{n}$ rate.[3]

In order to parametrically identify the ALR, we consider a specific parametric model for the the joint distributions of x, c and u. In particular, we use a correlated-random effects (CRE) model in which the relationship between c and x is modeled with a Mundlak (1978)-type specification.[4]

Assumption 2.1 *(Correlated Random Effects (CRE) Probit Model)* $\{x_i, u_i, c_i\}_{i=1}^n$ *are i.i.d. draws from the underlying population with the conditional distribution of* (u, c)*:*

$$c|x \sim N\left(\psi + \bar{x}\lambda, \sigma_c^2\right), \tag{5}$$

$$u|x, c \sim N(0, I_T), \tag{6}$$

where $\bar{x} = T^{-1}\sum_{t=1}^T x_t$*. For each* i*,* y_i *is generated according to the model in (1).*

Under Assumption 2.1, note that $M_j^t(x_t, x)$ can be written in terms of the model

[3] Altonji and Matzkin (2005) have a nonparametric version of ALR_j^t, which is a weighted version of the LAR with weights equal to one for each x_t.

[4] The Mundlak (1978)-type specification makes it particularly easy to deal with unbalanced panel data and also requires fewer parameters than the Chamberlain (1984) specification.

parameters as follows:

$$\begin{aligned} M_j^t(x_t, x) &= \int \theta_j^t(x_t, c) dH_c(c|x) \\ &= \frac{\partial \int m^t(x_t, c) dH_c(c|x)}{\partial x_{tj}} \\ &= \frac{\partial \Phi\big((x_t\beta + \psi + \bar{x}\lambda)/\sqrt{1+\sigma_c^2}\big)}{\partial x_{tj}} \\ &= \frac{\beta_j}{\sqrt{1+\sigma_c^2}} \phi\left(\frac{x_t\beta + \psi + \bar{x}\lambda}{\sqrt{1+\sigma_c^2}}\right) \\ &= \beta_{cj}\phi\left(x_t\beta_c + \psi_c + \bar{x}\lambda_c\right), \end{aligned}$$

where $\beta_c \equiv \frac{\beta}{\sqrt{1+\sigma_c^2}}$, $\psi_c \equiv \frac{\psi}{\sqrt{1+\sigma_c^2}}$, and $\lambda_c \equiv \frac{\lambda}{\sqrt{1+\sigma_c^2}}$ are the re-scaled parameters. Then, the ALR parameter is given by

$$\begin{aligned} ALR_j^t &= \int M_j^t(x_t, x) dG_x(x) \\ &= \int \beta_{cj}\phi\left(x_t\beta_c + \psi_c + \bar{x}\lambda_c\right) dG_x(x). \end{aligned} \tag{7}$$

If $(\hat{\beta}_c, \hat{\psi}_c, \hat{\lambda}_c)$ are consistent estimators for $(\beta_c, \psi_c, \lambda_c)$, then ALR can be consistently estimated by

$$\widehat{ALR}_j^t = \frac{1}{n}\sum_{i=1}^n \widehat{M}_j^t(x_{it}, x_i) = \frac{1}{n}\sum_{i=1}^n \hat{\beta}_{cj}\phi(x_{it}\hat{\beta}_c + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c). \tag{8}$$

The estimated partial effect $\widehat{ALR}_j^t$ is straightforward to obtain after estimation of the original model.

In order to formally prove the $\sqrt{n}$ convergence and asymptotic normality of the ALR estimator, we assume the following for the underlying estimators $(\hat{\beta}_c, \hat{\psi}_c, \hat{\lambda}_c)$:

Assumption 2.2 *Suppose $\hat{\beta}_c$, $\hat{\psi}_c$ and $\hat{\lambda}_c$ satisfy*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_c - \beta_c \\ \hat{\psi}_c - \psi_c \\ \hat{\lambda}_c - \lambda_c \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} \xi_\beta(y_i, x_i, \theta) \\ \xi_\psi(y_i, x_i, \theta) \\ \xi_\lambda(y_i, x_i, \theta) \end{pmatrix} + o_p(1) \tag{9}$$

where $\theta \equiv (\beta', \psi, \lambda', \sigma_c^2)$ and the ξ functions are measurable with finite second moments.

Assumption 2.2 implies that $(\hat{\beta}_c, \hat{\psi}_c, \hat{\lambda}_c)$ are asymptotically normally distributed by the

Central Limit Theorem. For example, the pooled probit estimator of y_t on $(x_t, 1, \bar{x})$ would satisfy this assumption — the expressions for the influence functions in that case are provided in the Appendix. There are obviously other estimators satisfying Assumption 2.2 for this model, but we will use pooled probit estimators since they are particularly easy to compute.

The following theorem states the asymptotic-normality result for $\widehat{ALR}_j^t$:

Theorem 2.3 *If Assumptions 2.1 and 2.2 hold,* $\sqrt{n}(\widehat{ALR}_j^t - ALR_j^t) \xrightarrow{D} N(0, \mathcal{V})$.

The exact expression for the asymptotic variance $\mathcal{V}$ is complicated and left for the Appendix. This asymptotic variance accounts for the first-stage estimation error. In general, different first-stage estimators will result in different asymptotic variances of $\widehat{ALR}_j^t$.

2.2 Conditional average local response (CALR)

The ALR of Section 2.1 is a natural measure of the "overall" effect of a covariate component x_{tj}, where the heterogeneity conditions upon the full covariate vector. A researcher may, however, be interested in the effect evaluated at a specific value for this component x_{tj}. This subsection introduces the *conditional average local response (CALR)* for this purpose, with the CALR being a function of x_{tj} along its support. The CALR, evaluated at x_{tj}, is defined as[5]

$$CALR_j^t(x_{tj}) \equiv E_x[M_j^t(x_t, x)|x_{tj}] = \int M_j^t(x_t, x) dG_x(x|x_{tj}). \tag{10}$$

This partial effect is the conditional expectation of LAR with respect to a single component of x_t. In the case of the CRE probit model above (Assumption 2.1), the CALR partial effect is

$$\begin{aligned} CALR_j^t(x_{tj}) &= \int M_j^t(x_t, x) dG_x(x|x_{tj}) \\ &= \int \beta_{cj} \phi\left(x_t \beta_c + \psi_c + \bar{x} \lambda_c\right) dG_x(x|x_{tj}), \end{aligned} \tag{11}$$

[5] Although we focus upon partial effects of the j-th component evaluated at values of the j-component, a similar expression could be written for the effect of the j-th component evaluated at values of the k-th component for $j \neq k$.

which can be estimated at a particular x_{tj}^* value by

$$\widehat{CALR}_j^t(x_{tj}^*) = \frac{\sum_{i=1}^n \widehat{M}_j^t(x_{it}, x_i) K\left(\frac{x_{itj} - x_{tj}^*}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_{itj} - x_{tj}^*}{h}\right)}, \tag{12}$$

where $K(\cdot)$ is a kernel function and h is the bandwidth. The following assumption on the kernel function and bandwidth is made:

Assumption 2.4

1. *The kernel function $K(u)$ is supported on $[-1, 1]$, symmetric around 0, and twice continuously differentiable.*

2. *The bandwidth h satisfies $\lim_{n\to\infty} nh^5 = 0$ and $\lim_{n\to\infty} nh = \infty$.*

The asymptotic distribution of the CALR estimator is then given by

Theorem 2.5 *If Assumptions 2.1, 2.2, and 2.4 hold, then*

$$\sqrt{nh}(\widehat{CALR}_j^t(x_{tj}^*) - CALR_j(x_{tj}^*)) \xrightarrow{D} N(0, \mathcal{V}_c). \tag{13}$$

Theorem 2.5 shows that the CALR estimator is asymptotical normal with a convergence rate that is the same as for single-dimensional nonparametric estimation (since matching is being done on a single covariate component). The expression for $\mathcal{V}_c$ is provided in the Appendix.[6]

2.3 Conditional average partial effect (CAPE)

The partial-effect idea first introduced by Chamberlain (1984), and extensively discussed by Wooldridge (2005b), will be called the *conditional average partial effect (CAPE)* in this paper. This effect is *conditional* in the sense that it is evaluated at a particular covariate vector x_t^0, although it will be unconditional with respect to the heterogeneity

[6] Unlike the ALR estimator, the estimation error from first-stage estimation can be ignored for the CALR. The first-stage estimation error is $O_p(n^{-1/2})$, which is of smaller order compared to the nonparametric convergence rate $O_p((nh)^{-1/2})$. Therefore, the choice of first-stage estimators will have no effect on the asymptotic variance of $\widehat{CALR}_j^t$.

distribution. Specifically, the CAPE with respect to component j for time t, evaluated at x_0^t, is defined as

$$CAPE_j^t(x_t^0) = E_c[\theta_j^t(x_t^0, c)]. \tag{14}$$

Note that, unlike the ALR and CALR partial-effects discussed previously, the integration is taken over the unconditional distribution of c without regard to the covariate value x_t^0.

For the CRE probit model (Assumption 2.1), the partial effect $CAPE_j^t(x_t^0)$ is

$$CAPE_j^t(x_t^0) = E_c[\theta_j^t(x_t^0, c)] = E_x\left[E_{c|x}[\theta_j^t(x_t^0, c)]\right] = E_x[\beta_{cj}\phi\left(x_t^0\beta_c + \psi_c + \bar{x}\lambda_c\right)],$$

which can be consistently estimated by

$$\widehat{CAPE}_j^t(x_t^0) = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{cj}\phi\left(x_t^0\hat{\beta}_c + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c\right). \tag{15}$$

Since the model parameters are estimated at a $\sqrt{n}$ rate, the CAPE partial-effect estimator in (15) is also clearly $\sqrt{n}$-consistent. Note that the ALR estimate in (8) and the CAPE estimate in (15) look very similar, with the important difference that x_{it} appears in ALR and the fixed x_t^0 appears in CAPE. For the former, the appearance of x_{it} in the same term as $\bar{x}_i$ (the heterogeneity control) entails using the conditional distribution $c_i|x_{it}$ in estimating the partial effects; for the latter, the use of the same x_t^0 for each i in the summation entails using the unconditional distribution of c_i.

A natural choice for x_t^0 may be the sample average $\bar{x}_t$, so that the estimated partial effect $\widehat{CAPE}_j^t(\bar{x}_t)$ estimates the underlying parameter $CAPE_j^t(E(x_t))$. If one wants to focus upon particular values of the component x_{tj}, we propose estimating the CAPE parameter at a specific value $x_{tj} = x_{tj}^*$ for that component and the conditional mean of the remaining components, that is $CAPE_j^t(x_{tj}^*, E(x_{t,-j}|x_{tj} = x_{tj}^*))$.[7] The subscript $_{-j}$ indicates all non-j components. This localized version of CAPE can be estimated by using

$$\bar{x}_{t,-j}(x_{tj}^*) = \frac{\sum_{i=1}^{n} x_{it,-j}K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)} \tag{16}$$

[7] This is a slight abuse of notation since CAPE is a function of a single vector argument. The vector argument is assumed to be composed of the two parts given by the arguments in the current notation.

and evaluating $\widehat{CAPE}_j^t(x_{tj}^*, \bar{x}_{t,-j}(x_{tj}^*))$.

2.4 Average partial effect (APE)

Finally, we consider a partial-effect concept that summarizes the CAPE effects but does not focus on a particular choice of x_t. This *average partial effect (APE)* is the population average, over x_t, of the CAPE effects:

$$APE_j^t = E_{x_t}[CAPE_j^t(x_t)]. \tag{17}$$

The average partial effect of component j at time t, again based upon the unconditional heterogeneity distribution, is given by

$$\widehat{APE}_j^t = \frac{1}{n}\sum_{i=1}^{n}\widehat{CAPE}_j^t(x_{it}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{i'=1}^{n}\hat{\beta}_{cj}\phi\left(x_{it}\hat{\beta}_c + \hat{\psi}_c + \bar{x}_{i'}\hat{\lambda}_c\right)\right). \tag{18}$$

The ALR and APE may be viewed as alternative summary measures for a covariate's partial effect, with the former (latter) using the conditional (unconditional) heterogeneity distribution.

2.5 CALR partial effects for discrete variables

Since the ALR-type partial effects considered in Sections 2.1 and 2.2 require use of the conditional heterogeneity distribution, it is non-trivial to directly extend the approach to discrete covariates. Altonji and Matzkin (2005), in a nonparametric context, consider only ALR effects for continuous covariates. If a researcher is interested in the partial effect of changing a binary covariate from zero to one, the question of whether to condition heterogeneity on the zero value or the one value arises. Such an issue is not relevant for APE or CAPE effects, as the heterogeneity distribution is unconditional and the continuous approach has a natural discrete-difference version (see, for example, Wooldridge (2010)).

We focus on the case of a binary covariate, but the approach can be generalized to any type of discrete covariate change. Suppose that x_{t1} (the first component of x_t) is a binary variable. For the partial effect of changing x_{t1} from 0 to 1, we focus on the subpopulation for which $x_{t1} = 0$ and the associated conditional heterogeneity distribution; fixing the conditional distribution, the effect is then estimated by shifting the covariate value. This CALR partial effect is denoted $CALR_{1,0\to1}^t$, and the counterpart to (3) in this case is $E_{u_t}[y_t|(1, x_{t,-1}), c] - E_{u_t}[y_t|(0, x_{t,-1}), c]$. The discrete version of the LAR is defined as

$$LAR^t_{1,0\to1} = \int \big(E_{u_t}[y_t|(1, x_{t,-1}), c] - E_{u_t}[y_t|(0, x_{t,-1}), c]\big) dH_c(c|x). \tag{19}$$

Under Assumption 2.1, $LAR^t_{1,0\to1}$ can be expressed

$$LAR^t_{1,0\to1} = \Phi\left(\beta_{c1} + x_{t,-1}\beta_{c,-1} + \psi_c + \bar{x}\lambda_c\right) - \Phi\left(x_{t,-1}\beta_{c,-1} + \psi_c + \bar{x}\lambda_c\right). \tag{20}$$

Then, $CALR^t_{1,0\to1}$ is given by

$$CALR^t_{1,0\to1} = \int LAR^t_{1,0\to1} dG_{x_{t,-1}}(x_{t,-1}|x_{t1} = 0), \tag{21}$$

which can be estimated by the sample analogue

$$\widehat{CALR}^t_{1,0\to1} = \frac{1}{n_{t0}} \sum_{i:x_{it1}=0} \left[\Phi\left(\hat{\beta}_{c1} + x_{it,-1}\hat{\beta}_{c,-1} + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c\right) - \Phi\left(x_{it,-1}\hat{\beta}_{c,-1} + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c\right)\right], \tag{22}$$

where n_{t0} denotes the number of observations with $x_{it1} = 0$.

The partial effect of a shift from 1 to 0, denoted $CALR^t_{1,1\to0}$, would be defined analogously (with 0 and 1 swapping roles). The effect would be estimated by

$$\widehat{CALR}^t_{1,1\to0} = \frac{1}{n_{t1}} \sum_{i:x_{it1}=1} \left[\Phi\left(x_{it,-1}\hat{\beta}_{c,-1} + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c\right) - \Phi\left(\hat{\beta}_{c1} + x_{it,-1}\hat{\beta}_{c,-1} + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c\right)\right], \tag{23}$$

where n_{t1} denotes the number of observations with $x_{it1} = 1$.

Note that a discrete-covariate version of the ALR effect can also be introduced since the quantity $LAR^t_{1,0\to1}$ is well-defined regardless of the value of x_{t1}. The ALR effect is defined as

$$ALR^t_{1,0\to1} = \int LAR^t_{1,0\to1} dG_{x_{t,-1}}(x_{t,-1}),$$

and can be estimated by

$$\widehat{ALR}^t_{1,0\to1} = \frac{1}{n}\sum_{i=1}^{n}\left[\Phi\left(\hat{\beta}_{c1} + x_{it,-1}\hat{\beta}_{c,-1} + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c\right) - \Phi\left(x_{it,-1}\hat{\beta}_{c,-1} + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c\right)\right]$$
$$= \frac{n_{t0}}{n}\widehat{CALR}^t_{1,0\to1} - \frac{n_{t1}}{n}\widehat{CALR}^t_{1,1\to0}.$$

The ALR effect of a shift from 1 to 0 is defined in the obvious way, and in this case we have $ALR^t_{1,1\to0} = -ALR^t_{1,0\to1}$.

3 A probit example and Monte Carlo simulations

This section considers a simple design for the CRE probit model in order to highlight the potential quantitative differences among the different partial-effect quantities discussed in Section 2. The section also provides Monte Carlo results that illustrate the finite-sample performance of the various partial-effects estimators.

The CRE probit design that we consider is the following:

Example 3.1 *Let $(y_1, y_2, x_1, x_2, c, u')$, where $u \equiv (u_1, u_2)'$, be generated as follows:*

$$y_t = 1(x_t + c + u_t < 0), \quad t \in \{1, 2\}$$
$$c|x_1, x_2 \sim a\bar{x},$$
$$u|x_1, x_2, c \sim N(0, I_2),$$
$$(x_1, x_2) \sim N(0, I_2).$$

where $a > 0$.

Note that this design has c has a deterministic function of x_1 and x_2. The scalar constant a dictates how important the heterogeneity is for the outcome y_t.

If $(y_1, y_2, x_1, x_2, c, u')$ are generated according to Example 3.1, then one can derive the following closed-form expressions for ALR, CALR, CAPE, and APE:

$$
\begin{aligned}
ALR &= \phi(0)\sqrt{\frac{2}{4+2a+a^2}} \\
CALR(x_1) &= \phi\left(\frac{2+a}{\sqrt{4+a^2}}x_1\right)\frac{2}{\sqrt{4+a^2}} \\
CAPE(x_1) &= \phi\left(\sqrt{\frac{2}{2+a^2}}x_1\right)\sqrt{\frac{2}{2+a^2}} \\
APE &= \phi(0)\sqrt{\frac{2}{4+a^2}}
\end{aligned}
$$

From the expressions above, it is clear that ALR can differ from APE (ALR $\leq$ APE with equality only when $a = 0$ (no heterogeneity)) and CALR can differ from CAPE (with the relative size depending on both a and the covariate value x_1).[8]

To illustrate the quantitative differences between the four different partial effects, we graph them (as a function of x_1) in Figures 1 and 2 corresponding to $a = 1$ and $a = 2$, respectively. The horizontal lines in the Figures represent the ALR and APE effects, neither of which is a function of x_1. The squares and circles on the CALR and CAPE curves, respectively, mark the effects evaluated at the $0.1, 0.2, \ldots, 0.9$ quantiles of x_1.[9] As mentioned above, the APE is larger than the ALR. The differences are *substantial* — 18.34% larger for $a = 1$, 22.47% larger for $a = 2$. For the CALR and CAPE effects, the difference in the curves becomes much more pronounced in the $a = 2$ case. That is, the difference between treating heterogeneity conditionally or unconditionally becomes larger when the heterogeneity is more related to the observed covariates.

Having shown how the true partial-effect parameters can differ from each other, we turn to estimation of the partial-effect parameters. Tables 2 and 3 summarize the results, again corresponding to the $a = 1$ and $a = 2$ designs, respectively. Results for ALR and APE are reported, as well as estimates of CALR and CAPE at the $0.1, 0.2, \ldots, 0.9$ quantiles of x_1. For CALR, the Epanechnikov kernel ($K(u) = 1(|u| < 1) \cdot 0.75(1 - u^2)$) is used with a bandwidth $h = 2\sigma_x n^{-1/4}$ where σ_x is the standard deviation of x_1. Sample sizes of $n = 250$ and $n = 1000$ are considered. In each row, the true parameter value is provided, and the bias and root-mean-squared-error for the estimators (based upon 10,000

[8] CALR and CAPE coincide when $a = 0$ (both equal to $\phi(x_1)$), but the fact that they coincide here is tied to the fact that x_1 and x_2 are independent in the design.

[9] Since x_t is scalar in this design, there is no need to account for $x_{t,-j}$ in the CAPE effect.

simulations) are reported. There is a minimal amount of bias across the various estimates. A comparison of the root MSE values for $n = 250$ and $n = 1000$ clearly demonstrate the $\sqrt{n}$-convergence rate of the underlying ALR, APE, and CAPE estimators.

4 Empirical application

We consider an empirical application in which we estimate the partial effect of doctor visits on health satisfaction. The data is taken from the first twelve annual waves (1984-1995) of the German Socioeconomic Panel (GSOEP). These data have been previously used in the literature by Wagner, Burkhauser, and Behringer (1993) and Riphahn, Wambach, and Million (2003).[10] To simplify matters, we considered a balanced panel data with $T = 2$ by taking the first two periods of data for any individual who appears more than twice in the dataset. The resulting dataset has 3,417 individuals and 6,834 observations.

The dependent variable of interest is the indicator variable *healthy*, defined to be equal to 1 if the individual has a self-reported health satisfaction greater than or equal to 7 (on a scale from 0 to 10).[11] The covariates used in the analysis include age, an indicator variable for being handicapped, monthly household net income, number of doctor visits in last three months, number of hospital visits in the last year, and an indicator variable for public health insurance. Table 4 provides descriptive statistics for the variables of interest in each of the two time periods.

Under the Mundlak specification of Assumption 2.1, the estimates from the pooled probit specification are reported in Table 5. The table provides the appropriately rescaled coefficient estimates for β and λ along with bootstrap standard errors (based upon bootstrap replications). As expected, the coefficient estimate for number of doctor visits is significantly negative (z-stat = 6.4). Those with more doctor visits in a given year are less likely to report being in good health.

Based upon the estimates in Table 5, we consider estimation of the partial effect of doctor visits upon health satisfaction. The estimates for ALR and APE are based upon equations (8) and (18), respectively. For the CALR and CAPE estimators, the bandwidth for $t = 1$ and $t = 2$ are chosen to be 1.270 and 1.277, respectively, based upon $1.06 * \hat{\sigma}^t n^{-1/5}$ where $\hat{\sigma}^t$ is the standard deviation of the $docvis_{it}$. Both CALR and CAPE

[10] The data are available from the Journal of Applied Econometrics Data Archive at http://www.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/.

[11] Different cutoffs for the binary-variable definition yield similar qualitative results.

are evaluated at number of doctor visits equal to $0, 1, \ldots, 10$. For CAPE, as suggested in Section 2.3, the estimates are calculated at the conditional average of the remaining covariates. The partial-effect results are summarized in Tables 6 and 7 for $t = 1$ and $t = 2$, respectively.

In line with the significant coefficient estimate from Table 5, the partial effect of doctor visits on health satisfaction is significantly negative for all the parameters that are estimated. The magnitude of the effects are slightly larger in the second time period, e.g. roughly 2% higher for APE. While the difference between ALR and APE estimates is not as large as seen in our Monte Carlo simulations, there are important differences between the comparable CALR and CAPE estimates. In both time periods and at all values (between 0 and 10) for the doctor-visits covariate, the CAPE estimate is larger than the CALR estimate. The final column in Tables 6 and 7 reports the ratio between the CAPE estimate and the corresponding CALR estimate. The CAPE estimates at $t = 1$ are at least 3% higher than the CALR estimates, with much larger differences seen at the extreme covariate values — 7% higher at 0, 11% higher at 10. Qualitatively similar results are found for $t = 2$, again with bigger differences between the CALR and CAPE estimates at the extreme values of doctors visits.

Appendix

Pooled probit model: Recall that $\theta' = (\beta', \psi, \lambda', \sigma_c^2)$. Define $\tilde{x}'_{it} = (x'_{it}, 1, \bar{x}'_i)$, $\theta'_c = (\beta'_c, \psi_c, \lambda'_c)$ and $\hat{\theta}'_c = (\hat{\beta}'_c, \hat{\psi}_c, \hat{\lambda}'_c)$. The pooled probit estimator is defined as

$$\hat{\theta}_c \equiv \arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T} \sum_{t=1}^{T} \Big(y_{it} \log \Phi(\tilde{x}_{it}\theta) + (1 - y_{it}) \log(1 - \Phi(\tilde{x}_{it}\theta)) \Big) \right).$$

Define $H_i(\theta_c)$ as

$$H(\theta_c) = E\left[\frac{1}{T} \sum_{t=1}^{T} \frac{\phi^2(\tilde{x}_{it}\theta_c)\tilde{x}_{it}\tilde{x}'_{it}}{\Phi(\tilde{x}_{it}\theta_c)(1 - \Phi(\tilde{x}_{it}\theta_c))} \right]$$

which is the expected Hessian for the maximization problem. Then, by standard M-estimation, we have

$$\begin{pmatrix} \xi_\beta(y_i, x_i, \theta) \\ \xi_\psi(y_i, x_i, \theta) \\ \xi_\lambda(y_i, x_i, \theta) \end{pmatrix} \equiv -H(\theta_c)^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}\frac{\phi(\tilde{x}_{it}\theta_c)\tilde{x}_{it}}{\Phi(\tilde{x}_{it}\theta_c)(1-\Phi(\tilde{x}_{it}\theta_c))}\Big).$$

Proof of Theorem 2.3: First, we define $\mathcal{A}(\beta_c, \psi_c, \lambda_c) \equiv E[\phi(x_{it}\beta_c + \psi_c + \bar{x}_i\lambda_c)]$. Hence, we define $\nabla\mathcal{A}(\beta_c, \psi_c, \lambda_c)' = E\left[\phi(x_{it}\beta_c + \psi_c + \bar{x}_i\lambda_c)(x_{it}', 1, \bar{x}_i')\right]$, the transpose of the gradient of $\mathcal{A}(\beta_c, \psi_c, \lambda_c)$ with respect to the re-scaled parameters. Note that

$$\begin{aligned}
&\sqrt{n}(\widehat{ALR}_j^t - ALR_j^t) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\beta}_{cj}\phi(x_{it}\hat{\beta}_c + \hat{\psi}_c + \bar{x}_i\hat{\lambda}_c) - ALR_j^t) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\beta_{cj}\phi(x_{it}\beta_c + \psi_c + \bar{x}_i\lambda_c) - ALR_j^t) + \left(\frac{1}{n}\sum_{i=1}^{n}\phi(x_{it}\beta_c + \psi_c + \bar{x}_i\lambda_c)\right)\sqrt{n}(\hat{\beta}_{cj} - \beta_{cj}) \\
&+ \left(\frac{\beta_{cj}}{n}\sum_{i=1}^{n}\phi(x_{it}\beta_c + \psi_c + \bar{x}_i\lambda_c)(x_{it}', 1, \bar{x}_i')\right)\sqrt{n}\begin{pmatrix} \hat{\beta}_c - \beta_c \\ \hat{\psi}_c - \psi_c \\ \hat{\lambda}_c - \lambda_c \end{pmatrix} + o_p(1) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg((\beta_{cj}\phi(x_{it}\beta_c + \psi_c + \bar{x}_i\lambda_c) - ALR_j^t) + \mathcal{A}(\beta_c, \psi_c, \lambda_c)\xi_{\beta_j}(y_i, x_i, \theta) \\
&+ \beta_{cj}\nabla\mathcal{A}(\beta_c, \psi_c, \lambda_c)'\begin{pmatrix} \xi_\beta(y_i, x_i, \theta) \\ \xi_\psi(y_i, x_i, \theta) \\ \xi_\lambda(y_i, x_i, \theta) \end{pmatrix}\Bigg) + o_p(1) \\
\equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{ALR_j^t}(y_i, x_i, \theta) + o_p(1).
\end{aligned}$$

The second equality holds from a Taylor expansion, and the third equality holds given the influence function representations of $\hat{\beta}_c$, $\hat{\psi}_c$, and $\hat{\lambda}_c$ and the fact that

$$\frac{1}{n}\sum_{i=1}^{n}\phi(x_{ti}\beta_c + \psi_c + \bar{x}_i\lambda_c) \xrightarrow{p} \mathcal{A}(\beta_c, \psi_c, \lambda_c)$$
$$\frac{1}{n}\sum_{i=1}^{n}\phi(x_{ti}\beta_c + \psi_c + \bar{x}_i\lambda_c)(x_{ti}', 1, \bar{x}_i') \xrightarrow{p} \nabla\mathcal{A}(\beta_c, \psi_c, \lambda_c)'.$$

Hence, we define $\xi_{ALR}(y_i, x_i, \theta)$ the influence function for $\widehat{ALR}_j^t$. And $\mathcal{V} = E[\xi_{ALR_j^t}(y_i, x_i, \theta)^2]$.

Proof of Theorem 2.5: First, by mean value expansion of $\widehat{M}_j^t(x_t, x)$ around the true scaled parameters and given the fact that the gradient of $M_j^t(x_t, x)$ with respect to the scaled parameters is bounded and that the estimators for scaled parameters are $\sqrt{n}$ consistent, then we have

$$\begin{aligned}
&\sup_{x_t,x}\left|\widehat{M}_j^t(x_t, x) - M_j^t(x_t, x)\right| \\
=&\sup_{x_t,x}\left|\hat{\beta}_{cj}\phi(x_t\hat{\beta}_c + \hat{\psi}_c + \bar{x}\hat{\lambda}_c) - \beta_{cj}\phi(x_t\beta_c + \psi_c + \bar{x}\lambda_c)\right| = O_p(n^{-1/2}).
\end{aligned}$$

This implies that

$$\sqrt{nh}\left|\frac{\sum_{i=1}^n \widehat{M}_j^t(x_{it}, x_i)K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)} - \frac{\sum_{i=1}^n M_j^t(x_{it}, x_i)K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)}\right| = o_p(1).$$

As a result, the CALR estimator based on $\widehat{M}_j^t(x_{it}, x_i)$ is asymptotically equivalent to the CALR estimator based on $M_j^t(x_{it}, x_i)$. Therefore, let $f_{x_{tj}}(x_{tj}^*)$ be the probability density function of x_{tk} evaluated at x_{tj}^* and $\sigma^2(x_{tj}^*) = E[M_j^t(x_{it}, x_i)^2|x_{tj}^*] - E[M_j^t(x_{it}, x_i)|x_{tj}^*]^2$ be the conditional variance of $M_j^t(x_{it}, x_i)$ evaluated at x_{tj}^*. Hence, by the theory for Nadaraya-Watson kernel estimator, we have $\mathcal{V}_c = \|K\|_2^2\sigma^2(x_{tj}^*)/f_{x_{tj}(x_{tj}^*)}$ where $\|K\|_2 = \sqrt{\int_u K^2(u)du}$ and $\sqrt{nh}(\widehat{CALR}_j^t(x_{tj}^*) - CALR_j^t(x_{tj}^*)) \xrightarrow{D} N(0, \mathcal{V}_c)$. Furthermore, $\mathcal{V}_c$ can be consistently estimated by $\widehat{\mathcal{V}}_c = \|K\|_2^2\hat{\sigma}^2(x_{tj}^*)/\hat{f}_{x_{tj}}(x_{tj}^*)$ where

$$\begin{aligned}
\hat{f}_{x_{tj}}(x_{tj}^*) &= \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_{itj} - x_{tj}^*}{h}\right), \\
\hat{\sigma}^2(x_{tj}^*) &= \frac{\sum_{i=1}^n \left(\widehat{M}_j^t(x_{it}, x_i) - \widehat{CALR}(x_{tj}^*)\right)^2 K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_{itj}-x_{tj}^*}{h}\right)}.
\end{aligned}$$

References

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

Blundell, Richard W., and James L. Powell, 2003, Endogeneity in nonparametric and semiparametric regression models, in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress (Econometric Society Monographs)*, Volume 2, eds. M. Dewatripont, L. P. Hansen, S. J. Turnovsky. Cambridge University Press, pp. 312-357.

655-679.

Carro, Jesus M., 2007, Estimating dynamic panel data discrete choice models with fixed effects, *Journal of Econometrics*, 140: 503-528.

Chamberlain, Gary, 1984, Panel data, in *Handbook of Econometrics*, Volume 2, eds. Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland, pp. 1247-1318.

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

Riphahn, Regina T., Achim Wambach, and Andreas Million, 2003, Incentive effects in the demand for health care: a bivariate panel count data estimation, *Journal of Applied Econometrics*, 13: 387-405.

Wagner, Gert G., Richard V. Burkhauser and Frederike Behringer, 1993, The English language public use file of the German Socio-Economic Panel, *Journal of Human Resources*, 28: 429-433.

Wooldridge, Jeffrey M., 2005a, Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of Applied Econometrics*, 20: 39-54.

Wooldridge, Jeffrey M., 2005b, Unobserved heterogeneity and estimation of average partial effects, in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. K. Andrews and J. H. Stock. Cambridge: Cambridge University Press, pp. 27-55.

Wooldridge, Jeffrey M., 2010, *Econometric analysis of cross section and panel data*, MIT Press.

Figure 1: True partial effects for CRE probit design ($a = 1$)

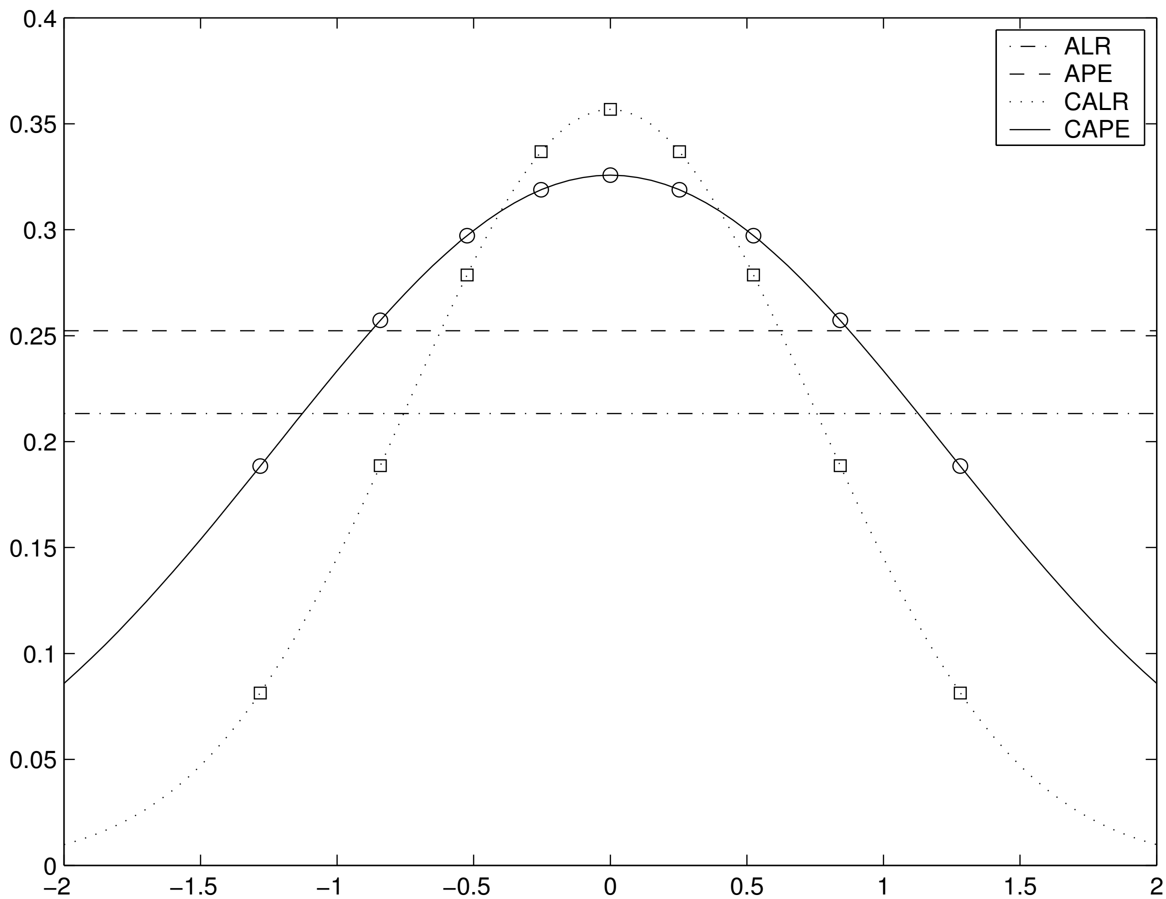

Figure 2: True partial effects for CRE probit design ($a = 2$)

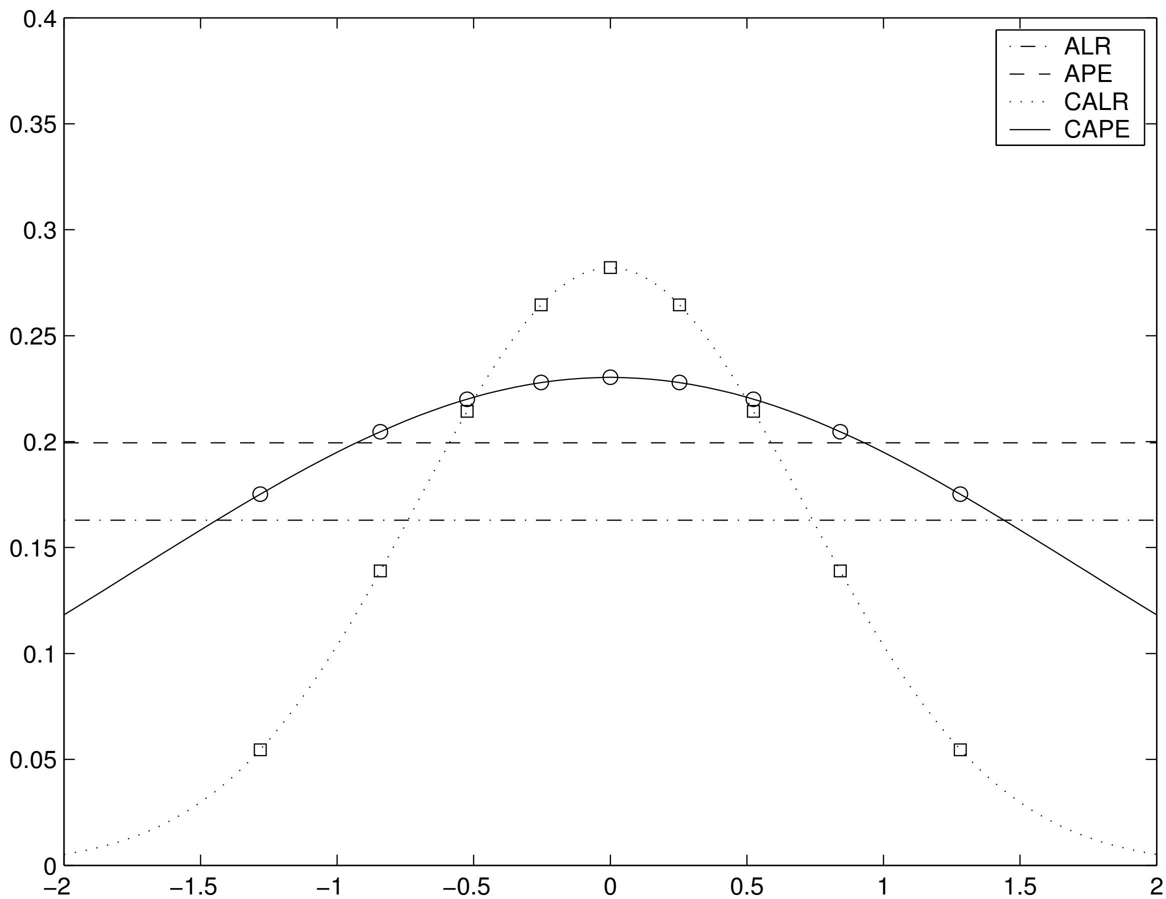

Table 2: Simulation results when $a = 1$.

		$n = 250$		$n = 1000$	
Parameter	True value	Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$
ALR	0.2132	0.0001	0.0224	0.0000	0.0110
APE	0.2523	-0.0007	0.0235	-0.0002	0.0116
CALR(q_1)	0.0814	0.0169	0.0254	0.0091	0.0135
CALR(q_2)	0.1886	0.0118	0.0272	0.0065	0.0144
CALR(q_3)	0.2786	0.0006	0.0341	0.0001	0.0173
CALR(q_4)	0.3368	-0.0088	0.0437	-0.0050	0.0223
CALR(q_5)	0.3568	-0.0124	0.0478	-0.0071	0.0246
CALR(q_6)	0.3368	-0.0089	0.0437	-0.0051	0.0223
CALR(q_7)	0.2786	0.0002	0.0340	0.0002	0.0171
CALR(q_8)	0.1886	0.0114	0.0271	0.0064	0.0139
CALR(q_9)	0.0814	0.0167	0.0253	0.0091	0.0134
CAPE(q_1)	0.1884	-0.0043	0.0163	-0.0010	0.0074
CAPE(q_2)	0.2572	-0.0021	0.0238	-0.0005	0.0120
CAPE(q_3)	0.2972	0.0006	0.0374	0.0001	0.0187
CAPE(q_4)	0.3188	0.0025	0.0467	0.0005	0.0229
CAPE(q_5)	0.3257	0.0031	0.0498	0.0007	0.0243
CAPE(q_6)	0.3188	0.0024	0.0467	0.0005	0.0228
CAPE(q_7)	0.2972	0.0004	0.0375	0.0001	0.0185
CAPE(q_8)	0.2572	-0.0023	0.0239	-0.0005	0.0118
CAPE(q_9)	0.1884	-0.0046	0.0163	-0.0011	0.0074

Note: q_j denotes the $0.j$ quantiles of x_1.

Table 3: Simulation results when $a = 2$.

		$n = 250$		$n = 1000$	
Parameter	True value	Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$
ALR	0.1629	0.0001	0.0207	-0.0002	0.0103
APE	0.1995	-0.0001	0.0252	-0.0004	0.0126
CALR(q_1)	0.0546	0.0135	0.0224	0.0070	0.0120
CALR(q_2)	0.1389	0.0106	0.0261	0.0055	0.0137
CALR(q_3)	0.2143	0.0010	0.0323	0.0000	0.0167
CALR(q_4)	0.2646	-0.0075	0.0405	-0.0049	0.0211
CALR(q_5)	0.2821	-0.0109	0.0441	-0.0068	0.0232
CALR(q_6)	0.2646	-0.0078	0.0407	-0.0048	0.0211
CALR(q_7)	0.2143	0.0005	0.0323	-0.0001	0.0165
CALR(q_8)	0.1389	0.0103	0.0260	0.0054	0.0138
CALR(q_9)	0.0546	0.0137	0.0227	0.0071	0.0121
CAPE(q_1)	0.1752	-0.0022	0.0174	-0.0008	0.0086
CAPE(q_2)	0.2047	-0.0001	0.0275	-0.0005	0.0137
CAPE(q_3)	0.2200	0.0013	0.0342	-0.0002	0.0169
CAPE(q_4)	0.2279	0.0021	0.0380	-0.0001	0.0187
CAPE(q_5)	0.2303	0.0024	0.0392	-0.0001	0.0192
CAPE(q_6)	0.2279	0.0021	0.0379	-0.0000	0.0186
CAPE(q_7)	0.2200	0.0012	0.0340	-0.0001	0.0168
CAPE(q_8)	0.2047	-0.0002	0.0272	-0.0004	0.0136
CAPE(q_9)	0.1752	-0.0023	0.0172	-0.0007	0.0084

Note: q_j denotes the $0.j$ quantiles of x_1.

Table 4: Descriptive statistics, German Socioeconomic Panel (GSOEP) data

	$t = 1$				$t = 2$			
variable	mean	std	min	max	mean	std	min	max
Health satisfaction (y)	0.634	0.482	0	1	0.618	0.486	0	1
Age	40.924	11.462	25	63	42.440	11.286	26	64
Handicapped	0.147	0.353	0	1	0.209	0.407	0	1
Household net income (monthly)	0.308	0.152	0.03	3	0.327	0.158	0	1.6
Doctor visits (x_{tj})	3.129	6.099	0	100	3.351	6.136	0	90
Hospital visits	0.145	0.911	0	30	0.152	1.306	0	51
Public health insurance	0.889	0.314	0	1	0.887	0.316	0	1

Table 5: Pooled probit estimates (re-scaled parameters), GSOEP data

Variable	Estimate	Bootstrap s.e.
β estimates:		
Age	-0.033	0.015
Handicapped	0.026	0.066
Household net income	-0.004	0.259
Doctor visits	-0.036	0.007
Hospital visits	-0.057	0.032
Public health insurance	-0.225	0.140
λ estimates (Mundlak specification):		
Age (mean)	0.010	0.015
Handicapped (mean)	-0.497	0.086
Household net income (mean)	0.396	0.300
Doctor visits (mean)	-0.051	0.008
Hospital visits (mean)	0.029	0.046
Public health insurance (mean)	-0.063	0.157

Table 6: Estimated partial effects for $t = 1$, GSOEP data

parameter	estimate	std	95% CI	CAPE/CALR ratio
ALR	-0.0106	0.0022	[-0.0145, -0.0058]	
APE	-0.0107	0.0022	[-0.0148, -0.0065]	
CALR(0)	-0.0101	0.0021	[-0.0138, -0.0056]	
CALR(1)	-0.0104	0.0022	[-0.0143, -0.0057]	
CALR(2)	-0.0111	0.0024	[-0.0153, -0.0060]	
CALR(3)	-0.0114	0.0025	[-0.0157, -0.0061]	
CALR(4)	-0.0116	0.0025	[-0.0160, -0.0062]	
CALR(5)	-0.0118	0.0025	[-0.0164, -0.0063]	
CALR(6)	-0.0118	0.0025	[-0.0165, -0.0063]	
CALR(7)	-0.0118	0.0026	[-0.0165, -0.0064]	
CALR(8)	-0.0117	0.0026	[-0.0162, -0.0064]	
CALR(9)	-0.0116	0.0024	[-0.0161, -0.0063]	
CALR(10)	-0.0112	0.0024	[-0.0155, -0.0061]	
CAPE(0)	-0.0108	0.0022	[-0.0147, -0.0059]	1.07
CAPE(1)	-0.0110	0.0023	[-0.0150, -0.0060]	1.06
CAPE(2)	-0.0115	0.0024	[-0.0158, -0.0061]	1.04
CAPE(3)	-0.0117	0.0025	[-0.0161, -0.0062]	1.03
CAPE(4)	-0.0119	0.0025	[-0.0164, -0.0063]	1.03
CAPE(5)	-0.0121	0.0026	[-0.0167, -0.0064]	1.03
CAPE(6)	-0.0122	0.0026	[-0.0169, -0.0064]	1.03
CAPE(7)	-0.0123	0.0027	[-0.0170, -0.0065]	1.04
CAPE(8)	-0.0124	0.0027	[-0.0170, -0.0066]	1.06
CAPE(9)	-0.0124	0.0027	[-0.0170, -0.0065]	1.07
CAPE(10)	-0.0124	0.0027	[-0.0170, -0.0065]	1.11

Note: CALR(j) and CAPE(j) are partial-effects parameters evaluated at number of doctor visits = j.

Table 7: Estimated partial effects for $t = 2$, GSOEP data

parameter	estimate	std	95% CI	CAPE/CALR ratio
ALR	-0.0107	0.0022	[-0.0147, -0.0058]	
APE	-0.0109	0.0021	[-0.0149, -0.0065]	
CALR(0)	-0.0103	0.0021	[-0.0141 , -0.0056]	
CALR(1)	-0.0107	0.0022	[-0.0146 , -0.0057]	
CALR(2)	-0.0113	0.0024	[-0.0156 , -0.0060]	
CALR(3)	-0.0116	0.0025	[-0.0160 , -0.0061]	
CALR(4)	-0.0117	0.0025	[-0.0162 , -0.0062]	
CALR(5)	-0.0118	0.0025	[-0.0163 , -0.0062]	
CALR(6)	-0.0118	0.0025	[-0.0163 , -0.0063]	
CALR(7)	-0.0120	0.0026	[-0.0166 , -0.0064]	
CALR(8)	-0.0122	0.0026	[-0.0169 , -0.0065]	
CALR(9)	-0.0120	0.0024	[-0.0160 , -0.0063]	
CALR(10)	-0.0114	0.0024	[-0.0157 , -0.0062]	
CAPE(0)	-0.0110	0.0023	[-0.0150, -0.0060]	1.07
CAPE(1)	-0.0112	0.0023	[-0.0153, -0.0061]	1.05
CAPE(2)	-0.0117	0.0025	[-0.0161, -0.0062]	1.04
CAPE(3)	-0.0119	0.0025	[-0.0164, -0.0063]	1.03
CAPE(4)	-0.0120	0.0026	[-0.0165, -0.0063]	1.03
CAPE(5)	-0.0121	0.0026	[-0.0167, -0.0064]	1.03
CAPE(6)	-0.0123	0.0026	[-0.0170, -0.0065]	1.04
CAPE(7)	-0.0123	0.0027	[-0.0171, -0.0065]	1.03
CAPE(8)	-0.0124	0.0027	[-0.0170, -0.0065]	1.02
CAPE(9)	-0.0124	0.0027	[-0.0170, -0.0065]	1.03
CAPE(10)	-0.0124	0.0027	[-0.0170, -0.0065]	1.09

Note: CALR(j) and CAPE(j) are partial-effects parameters evaluated at number of doctor visits = j.