Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models

Iván Fernández-Val
Department of Economics
Boston University

February 10, 2007

Abstract

Fixed effects estimators of nonlinear panel models can be severely biased due to the incidental parameters problem. In this paper I find that the most important component of this incidental parameters bias for probit fixed effects estimators of index coefficients is proportional to the true value of these coefficients, using a large-T expansion of the bias. This result allows me to derive a lower bound for this bias, and to show that fixed effects estimates of ratios of coefficients and average marginal effects have zero bias in the absence of heterogeneity, and have negligible bias relative to their true values for a wide variety of distributions of regressors and individual effects. New bias corrected estimators for index coefficients and marginal effects with improved finite sample properties are also proposed for static and dynamic probit, logit, and linear probability models with predetermined regressors.

JEL Classification: C23; C25; J22.

Keywords: Panel data; Bias; Discrete Choice Models; Probit; Fixed effects; Labor Force Participation.

Acknowledgment. I wish to thank Josh Angrist, Victor Chernozhukov, and Whitney Newey for encouragement and advice. I thank Jesús Carro for sharing his data. For suggestions and comments, I am grateful to Paula Bustos, Sylvain Chassang, Michael Jansson, Guido Kuersteiner, Arthur Lewbel, and the seminar participants at MIT, Chicago GSB, UBC, Harvard KSG, Boston University, Universidad de Navarra, LSE, CEMFI, UAB, UPF, Banco de España, University of Chicago, Brown University, Princeton University, and 2006 Winter Econometric Society Meetings. All remaining errors are mine. I gratefully acknowledge financial support from Fundación Caja Madrid and Fundación Ramón Areces. Please send comments or suggestions to ivanf@bu.edu.

1 Introduction

Panel data models are widely used in empirical economics because they allow researchers to control for unobserved individual time-invariant heterogeneity. However, these models pose important technical challenges in nonlinear and/or dynamic settings. In particular, if individual heterogeneity is left completely unrestricted, then estimates of model parameters suffer from the incidental parameters problem, first noted by Neyman and Scott (1948). This problem arises because the unobserved individual characteristics are replaced by inconsistent sample estimates, which, in turn, bias estimates of model parameters. Examples include probit with fixed effects, and linear and nonlinear models with lagged dependent variables and fixed effects (see, e.g., Nerlove, 1967; Nerlove, 1971; Heckman, 1981; Nickell, 1981; Greene, 2002; Katz, 2001; and Hahn and Newey, 2004).

In nonlinear models we often need to go beyond estimation of model parameters and obtain estimated marginal effects. In probit models, for example, the index coefficients cannot be interpreted as the effects of changes in the regressors on the conditional probability of the response. Accordingly, I deviate from previous studies of nonlinear panel models by looking at the properties of fixed effects estimators of marginal effects.[1] The motivation for this analysis comes from a question posed by Wooldridge: "How does treating the *individual effects* as parameters to estimate - in a "fixed effects probit" analysis - affect estimation of the APEs *(average partial effects)*?"[2] Wooldridge conjectures that the estimators of the marginal effects have reasonable properties. Here, using the expansion of the bias for the fixed effects estimators of index coefficients, I characterize the analytical expression for the bias of these average marginal effects. As Wooldridge anticipated, this bias is negligible relative to the true average effect for a wide variety of distributions of regressors and individual effects, and is identically zero in the absence of heterogeneity. This helps explain the small biases in the marginal effects estimates that Hahn and Newey (2004) (HN henceforth) find in Monte Carlo examples.

The derivation of the bias for the marginal effects estimators relies on the properties of the fixed effects estimators of model parameters and individual effects used in the construction of the marginal effects. A second contribution of the paper is to explore the bias properties of first

[1]Marginal effects are defined either as the change in the outcome conditional probability as a response to an one-unit increase in a regressor, or as a local approximation based on the slope of the outcome conditional probability. For example, in a probit with scalar regressor the marginal effects can be defined either as $\Phi((x+1)\theta) - \Phi(x\theta)$ or as $\theta\phi(x\theta)$, where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cdf and pdf of the standard normal distribution, respectively.

[2]C.f., Wooldridge (2002), p. 489 (italics mine).

effects estimators of model parameters (index coefficients) in probit models. A lower bound and a proportionality result are derived for this bias. The lower bound, which depends uniquely upon the number of time periods of the panel, establishes that the incidental parameters bias is at least 20 % for 4-period panels, and 10 % for 8-period panels. Proportionality, on the other hand, establishes that the bias of probit fixed effect estimators is a matrix-weighted average of the own parameters. This implies, for instance, that probit fixed effects estimates are biased away from zero when the regressor is scalar, providing a theoretical foundation for previous numerical evidence (see, for e.g., Greene, 2002). In the special case where there is no heterogeneity the matrices in the weighted average reduce to scalar multiples of the identity matrix. This suggests that fixed effects estimators of ratios of coefficients do not suffer from the incidental parameters bias in probit models when the level of individual heterogeneity is moderate. These ratios are often structural parameters of interest because they can be interpreted as marginal rates of substitution in many economic applications where the index coefficients are only identified up to scale.

Most of the theoretical results of the paper are concerned with static probit models with exogenous regressors, but I also explore related questions in other linear and nonlinear models with predetermined regressors and fixed effects. In particular, I find numerical evidence suggesting that probit and logit fixed effects estimates are biased downward for lagged dependent variables. This finding for marginal effects in dynamic nonlinear models resemble the result for fixed effects estimators of model parameters in dynamic linear models. I also develop new bias correction methods for estimates of index coefficients and marginal effects that exhibit better finite sample properties than the existing alternatives for probit and logit static and dynamic models. The improvement comes from exploiting more intensively the parametric structure of the problem by using expected quantities instead of observed quantities in the estimation of the bias.[3] Simple linear probability models, in the spirit of Angrist (2001), also perform well in estimating average marginal effects for exogenous regressors, but need to be corrected when the regressors are just predetermined.

The properties of probit and logit fixed effects estimators of model parameters and marginal effects are illustrated with an analysis of female labor force participation using 10 waves from the Panel Survey of Income Dynamics (PSID). The analysis here is motivated by similar studies in labor economics, where panel binary choice processes have been widely used to model

[3]This approach is similar to the use of the conditional information matrix in the estimation of asymptotic variances in maximum likelihood, instead of other alternatives, such as the sample average of the outer product of the scores or the sample average of the negative Hessian (Porter, 2002).

female labor force participation decisions (see, e.g., Hyslop, 1999; Chay and Hyslop, 2000; and Carro, 2006). I find that fixed effects estimators, while biased for index coefficients, give very similar estimates to their bias corrected counterparts for marginal effects in static models. On the other hand, uncorrected fixed effects estimators are biased for both index coefficients and marginal effects in dynamic models that account for true state dependence. In this case, the bias corrections presented here are effective reducing the incidental parameters problem.

The approach followed in this paper is related to the recent large-n large-T literature for panel data estimators, see, e.g., Lancaster (2000), Hahn and Kuersteiner (2002), Woutersen (2002), Arellano (2003), Alvarez and Arellano (2003), Hahn and Kuersteiner (2003), Hahn and Newey (2004), and Carro (2006). These studies aim to provide what I refer to as large-T consistent estimates because they rely on an asymptotic approximation to the behavior of the estimator that lets both the number of individuals, n, and the time dimension, T, grow with the sample size.[4] The idea behind these methods is to expand the incidental parameters bias of the estimator in orders of magnitude of T, and to remove an estimate of the leading term of the bias from the estimator.[5] As a result, the adjusted estimator has bias of order T^{-2}, whereas the bias of the initial estimator is T^{-1}. All these paper, however, focus mostly on estimation of model parameters. The only exception is Hahn and Newey (2004) that extends the bias corrections to marginal effects, and reports numerical evidence on the small bias of uncorrected fixed effects probit estimates of marginal effects (see also Carro, 2006). The results on this paper provide a theoretical explanation to these numerical findings.

The paper is organized as follows. Section 2 describes the panel binary choice model and the parameters of interest in this model. Section 3 reviews the incidental parameter problem for fixed effects estimators of model parameters, and characterizes the expression of the bias for the probit. Section 4 analyzes the properties of fixed effects estimators of marginal effects and establishes the small bias property for the probit. Monte Carlo results and the empirical illustration are given in Sections 5 and 6, respectively. Section 7 concludes. The proofs of the main results are given in the Appendix.

[4] Fixed-T-consistent estimators have also been derived for panel logit models (see Cox, 1958, Rasch, 1960, Andersen, 1973, Chamberlain, 1980, for the static case; and Cox, 1958, Chamberlain, 1985, and Honoré and Kyriazidou, 2000, for the dynamic case), and other semiparametric index models (see Manski, 1987, for the static case; and Honoré and Kyriazidou, 2000, for the dynamic case). These methods, however, do not provide estimates for individual effects, precluding estimation of other quantities of interest, such as marginal effects.

[5] To avoid complicated terminology, in the future I will generally refer to the first term of the large-T expansion of the bias simply as the bias.

2 The Model and Parameters of Interest

2.1 The Model

Given a binary response Y and a $p \times 1$ regressor vector X, the response for individual i at time t is assumed to be generated by the following single index process

$$Y_{it} = \mathbf{1}\left\{X_{it}'\theta_0 + \alpha_i - \epsilon_{it} \geq 0\right\}, \tag{2.1}$$

where $\mathbf{1}\{C\}$ is an indicator function that takes on value one if condition C is satisfied and zero otherwise; θ_0 denotes a $p \times 1$ vector of parameters (index coefficients); α_i is a scalar unobserved individual effect; and ϵ_{it} is a time-individual specific random shock. This is an error-components model where the unobserved error term is decomposed into a permanent individual-specific component α_i and a transitory shock ϵ_{it}. Examples of economic decisions that can be modeled within this framework include labor force participation, union membership, migration, purchase of durable goods, marital status, or fertility (see Amemiya, 1981, for a survey).

2.2 Structural Parameters: Index Coefficients

In binary choice models the conditional probability of the response, given the regressors and individual effects, is determined by the distribution of the random shock. Thus, we have

$$E[Y_{it}|X_{it}, \alpha_i] = \Pr[Y_{it}|X_{it}, \alpha_i] = F_\epsilon(X_{it}'\theta_0 + \alpha_i|X_{it}, \alpha_i), \tag{2.2}$$

where $F_\epsilon(\cdot|\cdot)$ is the conditional cdf of ϵ. Since F_ϵ is nondecreasing, the sign of the effect of a change on the regressor j, X_{jit}, is determined by the corresponding element of the vector of index coefficients $\theta_{j,0}$. Moreover, since the derivative of the conditional probability with respect to X_{jit} takes the form

$$\frac{\partial \Pr[Y_{it}|X_{it}, \alpha_i]}{\partial X_{jit}} = \theta_{j,0} f_\epsilon(X_{it}'\theta_0 + \alpha_i|X_{it}, \alpha_i), \tag{2.3}$$

where f_ϵ is the derivative of F_ϵ, ratios of two index coefficients give relative effects of small changes on the corresponding regressors. In random utility models, where the index process is the solution to an economic optimization program, these ratios of index coefficients can be interpreted as behavioral parameters (see, e.g., McFadden, 1974).

An important case of the previous model is when the regressor vector includes a lag of the response variable, i.e.,

$$Y_{it} = \mathbf{1}\left\{\theta_{y,0} Y_{i,t-1} + X_{it}'\theta_{x,0} + \alpha_i - \epsilon_{it} \geq 0\right\}. \tag{2.4}$$

In this dynamic specification persistence in the outcome can be a consequence of higher unobserved individual propensity to experience the event in all the periods, as measured by α_i, or to alterations in the individual behavior for having experienced the event in the previous period, as measured by $\theta_{y,0}$. Heckman (1981a) refers to these sources of persistence as heterogeneity and true state dependence, respectively, and argues about the importance of distinguishing between them because they have very different economic and policy implications. Examples of empirical studies that use this type of specification include Card and Sullivan (1988), Moon and Stotsky (1993), Roberts and Tybout (1997), Hyslop (1999), Chay and Hyslop (2000), and Carro (2006).

2.3 Marginal Effects

In empirical analysis the ultimate quantities of interest are often the marginal effects of specific changes in the regressors on the response conditional probability (see, e.g., Angrist, 2001; Ruud 2001; Wooldridge, 2002; and Wooldridge, 2005). Structural parameters in binary choice models, however, provide information about the sign and relative magnitude of these effects but not about their absolute magnitude. Thus, for example, for a model with two regressors, say X_1 and X_2, and corresponding parameters θ_1 and θ_2, the marginal effect of a one-unit increase in X_1 on the conditional probability of Y for individual i at time t, evaluated at $x_{it} = (x_{1it}, x_{2it})$, is given by

$$m(x_{it}, \theta, \alpha_i) := F_\epsilon((x_{1it} + 1)\theta_1 + x_{2it}\theta_2 + \alpha_i | X_{it}, \alpha_i) - F_\epsilon(x_{1it}\theta_1 + x_{2it}\theta_2 + \alpha_i | X_{it}, \alpha_i). \quad (2.5)$$

When X_1 is continuous, the previous expression is usually approximated by a local version based on the derivative of the conditional probability with respect to X_1, that is

$$\tilde{m}(x_{it}, \theta, \alpha_i) := \frac{\partial}{\partial x_{1it}} F_\epsilon(x_{1it}\theta_1 + x_{2it}\theta_2 + \alpha_i | X_{it}, \alpha_i) = \theta_1 f_\epsilon(x_{1it}\theta_1 + x_{2it}\theta_2 + \alpha_i | X_{it}, \alpha_i). \quad (2.6)$$

Marginal effects in nonlinear models depend on the individual effects α_i and the level chosen for evaluating the regressors x_{it}. This heterogeneity is in fact an attractive feature of these models because it allows, for instance, the marginal effects to be decreasing in the propensity (measured by the individual effect) to experience the event. Thus, individuals more prone to work are arguably less sensitive to marginal changes on other observable characteristics when deciding labor force participation. This heterogeneity, however, also raises the question of what are the relevant effects to report. A common practice is to give some summary measure, for example, the average observed effect or the effect for some interesting value of the regressors.

Chamberlain (1984) suggests reporting the average effect for an individual randomly drawn from the population, that is

$$\mu(x_1, \theta) = \int m((x_1, X_{2it}), \theta, \alpha_i) dG_{X_{2it}, \alpha_i}(X_{2it}, \alpha_i), \tag{2.7}$$

or

$$\mu(\theta) = \int \tilde{m}(X_{it}, \theta, \alpha_i) dH_{X_{it}, \alpha_i}(X_{it}, \alpha_i), \tag{2.8}$$

where G and H are the joint distribution of (X_{2it}, α_i) and (X_{it}, α_i), respectively, and x_1 is some interesting value of X_1. The previous measures correspond in fact to different thought experiments. The first measure, commonly used for discrete variables, corresponds to the counterfactual experiment where the change on the outcome probability is evaluated as if all the individuals would have chosen x_1 initially and receive an additional unit of X_1. The second measure, usually employed for continuous variables, is the average derivative of the response probabilities with respect to x_1, i.e., the average effect of giving one additional unit of X_1. These effects can be calculated also for subpopulations of interest by conditioning on the relevant values of the covariates. For example, if X_1 is binary (treatment indicator), the average treatment effect on the treated (ATT) is

$$\mu_1(\theta) = \int m((0, X_{2it}), \theta, \alpha_i) dG_{X_{2it}, \alpha_i}(X_{2it}, \alpha_i | X_{1it} = 1). \tag{2.9}$$

Other alternative measures used in cross-section models, such as the effect evaluated for an individual with average characteristics, are less attractive for panel data models because they raise conceptual and implementation problems (see Carro, 2006, for a related discussion about other measures of marginal effects). On the conceptual side, Chamberlain (1984) and Ruud (2000) argue that this effect may not be relevant for most of the population. The practical obstacle relates to the difficulty of estimating average characteristics in panel models. Thus, replacing population expectations for sample analogs does not always work in binary choice models estimated using a fixed-effects approach. The problem here is that the standard logit and probit estimates of the individual effects are unbounded for individuals that do not change status in the sample, and therefore the sample average of the estimated individual effects is generally not well defined.

It should be noted here that the previous definitions of marginal effects are static in the sense that they do not account for possible feedback from present changes on the response variable to future values of the regressors. Thus, for example, if the regressor includes lags of the dependent variable, the marginal effects considered only capture the instantaneous impact of a change in the regressor on the conditional probability of the response without accounting for the future changes in conditional probabilities through the lag response. Following Chamberlain interpretation, I measure the instantaneous (contemporaneous) effect of a change in a regressor on the conditional probability of the response for an individual randomly chosen from the population at a random period. Calculation of dynamic marginal effects requires imposing structure on the feedback process from the dependent variable to the regressors (see, e.g., Carneiro, Hansen, and Heckman, 2003).

3 Fixed Effects Estimation of Structural Parameters

3.1 Fixed Effects Maximum Likelihood Estimator

In economic applications regressors and individual heterogeneity are usually correlated because regressors are choice variables and individual heterogeneity usually represents variation in tastes or technology. To avoid imposing any structure on this relationship, I adopt a fixed-effects approach and treat the sample realization of the individual effects $\{\alpha_i\}_{i=1,\dots,n}$ as parameters to be estimated; see Mundlak (1978), Lancaster (2000), Arellano and Honoré (2001), and Arellano (2003) for a similar interpretation of fixed effects estimators. This approach has the additional advantage in dynamic models of not imposing restrictions on the initial values of the process avoiding the so-called initial condition problems (Heckman, 1981b).

To estimate the model parameters, we draw a sample of the observable variables $\{Y_{it}, X_{it}\}$, $t = 1, ..., T$, $i = 1, .., n$, where i and t usually index individuals and time periods, respectively.[6] Then, assuming that ϵ follows a known distribution conditional on regressors and individual effects, typically normal or logistic, a natural way of estimating this model is by maximum likelihood.[7] Thus, if ϵ_{it}'s are i.i.d. conditional on X_i^t and α_i, with cdf $F_\epsilon(\cdot|X_i^t, \alpha_i)$, the conditional log-likelihood for observation i at time t is

$$l_{it}(\theta, \alpha_i) := Y_{it} \log F_{it}(\theta, \alpha_i) + (1 - Y_{it}) \log(1 - F_{it}(\theta, \alpha_i)), \tag{3.1}$$

where $F_{it}(\theta, \alpha_i)$ denotes $F_\epsilon(X_{it}'\theta + \alpha_i | X_i^t, \alpha_i)$, and the maximum likelihood estimator (MLE) of

[6] In the sequel, for any random variable Z, Z_{it} denotes observation at period t for individual i; Z_i^t denotes a vector with all the observations for individual i until period t, i.e., $Z_i^t := \{Z_{i1}, ..., Z_{it}\}$.

[7] Since the inference is conditional on the realization of the regressors and individual effects, all the probability statements should be qualified with a.s. I omit this qualifier for notational convenience.

θ, concentrating out the α_i's, is the solution to

$$\widehat{\theta} := \arg\max_{\theta} \sum_{i=1}^{n} \sum_{t=1}^{T} l_{it}(\theta, \widehat{\alpha}_i(\theta))/nT, \quad \widehat{\alpha}_i(\theta) := \arg\max_{\alpha} \sum_{t=1}^{T} l_{it}(\theta, \alpha)/T. \tag{3.2}$$

It should be emphasized here that, unlike the common practice in fixed effects estimation of nonlinear panel models, I only assume that the model is dynamically complete, and that the regressor vector X is predetermined, instead of (strictly) exogenous. This is an important departure in the modeling assumptions as it permits, for instance, capturing richer dynamic feedbacks from the dependent variable to the regressors. The leading cases of predetermined variables are lagged dependent variables, but are not confined to them. Arellano and Carrasco (2003), for example, argue about the implausibility of the exogeneity assumption of fertility variables on the female labor participation decision and propose a random effects estimator that allows the regressors to be predetermined. Wooldridge (2001) derives a conditional maximum likelihood estimator by imposing a parametric assumption on the conditional distribution of the individual heterogeneity given the initial values of the predetermined regressors that is consistent under fixed-T asymptotics. Honoré and Lewbel (2002) and Lewbel (2005) propose alternative fixed-T consistent fixed effects estimators for binary choice models with predetermined regressors for the case where one of the regressors is exogenous and independent of the individual effects (conditional on the other regressors). Carro (2006) proposes a large-T bias corrected estimator for dynamic binary choice models that allows for lags of the dependent variable as regressors, but assumes that the rest of the explanatory variables are exogenous. In Section 5 I introduce large-T bias corrected fixed effects estimators for static and dynamic binary choice models with general predetermined regressors that do not impose any parametric assumptions on the individual heterogeneity.

3.2 Incidental Parameters Problem

Fixed effects MLEs generally suffer from the incidental parameters problem noted by Neyman and Scott (1948). This problem arises because the unobserved individual effects are replaced by sample estimates. In nonlinear and/or dynamic models, estimation of the model parameters cannot generally be separated from the estimation of the individual effects. Then, the estimation error of the individual effects contaminates the estimates of model parameters. To see this, for $Z_{it} := (Y_{it}, X_{it})$ and any function $m(Z_{it}, \alpha_i)$, let $E_n[m(Z_{it}, \alpha_i)] := \lim_{n\to\infty} \sum_{i=1}^{n} m(Z_{it}, \alpha_i)/n$, whenever this limit exists. Then, from the usual maximum likelihood properties, for $n \to \infty$

with T fixed,

$$\widehat{\theta} \xrightarrow{p} \theta_T, \;\; \theta_T := \arg\max_{\theta} E_n \left[\sum_{t=1}^{T} l_{it}(\theta, \widehat{\alpha}_i(\theta))/T \right]. \tag{3.3}$$

When the true conditional log-likelihood of Y is $l_{it}(\theta_0, \alpha_i)$ generally $\theta_T \neq \theta_0$ since $\widehat{\alpha}_i(\theta_0) \neq \alpha_i$; but $\theta_T \to \theta_0$ as $T \to \infty$, since $\widehat{\alpha}_i(\theta_0) \to \alpha_i$ as $T \to \infty$.

For the smooth likelihoods considered here, $\theta_T = \theta_0 + \frac{\mathcal{B}}{T} + O\left(\frac{1}{T^2}\right)$ for some $\mathcal{B}$. By asymptotic normality of the MLE, $\sqrt{nT}(\widehat{\theta} - \theta_T) \xrightarrow{d} N(0, -\mathcal{J}^{-1})$ as $n \to \infty$, and therefore

$$\sqrt{nT}(\widehat{\theta} - \theta_0) = \sqrt{nT}(\widehat{\theta} - \theta_T) + \sqrt{nT}\frac{\mathcal{B}}{T} + O_p\left(\sqrt{\frac{n}{T^3}}\right). \tag{3.4}$$

Here we can see that even if we let T grow at the same rate as n, that is $T = O(n)$, the MLE, while consistent, has a limiting distribution not centered around the true parameter value. Under asymptotic sequences where T grows large no faster than n, the estimates of the individual effects converge to their true values at a slower rate than the sample size nT, because only observations for each individual convey information about the corresponding individual effect. This slower rate translates into bias in the asymptotic distribution of the estimators of the model parameters. The presence of bias in the asymptotic distribution is the large-T version of the incidental parameters problem, which invalidates any inference based on this distribution.

3.3 Large-T Approximation to the Bias

Hahn and Kuersteiner (2003) characterize the expression for the bias using a stochastic expansion of the fixed effects estimator in orders of T for general dynamic models. Here I briefly reproduce this expansion and the resulting expression for the bias to introduce the notation that will be extensively used in the rest of the analysis. Let

$$u_{it}(\theta, \alpha) := \frac{\partial}{\partial \theta} l_{it}(\theta, \alpha), \quad v_{it}(\theta, \alpha) := \frac{\partial}{\partial \alpha} l_{it}(\theta, \alpha), \tag{3.5}$$

additional subscripts denote partial derivatives, e.g. $u_{it\theta}(\theta, \alpha) := \partial u_{it}(\theta, \alpha)/\partial \theta'$, and the arguments are omitted when the expressions are evaluated at the true parameter value, i.e. $v_{it\theta} = v_{it\theta}(\theta_0, \alpha_i)$. The (first term of the large-T expansion of the) asymptotic bias is

$$T(\widehat{\theta} - \theta_0) \xrightarrow{p} T(\theta_T - \theta_0) = -\mathcal{J}^{-1} b := \mathcal{B}, \tag{3.6}$$

where $\mathcal{J}$ is the probability limit of the Jacobian of the estimating equation for θ,

$$\mathcal{J} := E_n \left[E_T \left[u_{it\theta} \right] - E_T \left[u_{it\alpha} \right] \frac{E_T \left[v_{it\theta} \right]}{E_T \left[v_{it\alpha} \right]} \right], \tag{3.7}$$

where $E_T[h_{it}] := \lim_{T\to\infty} \sum_{t=1}^{T} h_{it}/T$; and b is the bias of the estimating equation for θ,

$$b := E_n \left\{ E_T[u_{it\alpha}]\,\beta_i + \bar{E}_T[u_{it\alpha}\psi_{is}] + \frac{1}{2}\sigma_i^2 E_T[u_{it\alpha\alpha}] \right\}, \tag{3.8}$$

where $\bar{E}_T[h_{it}k_{is}] := \sum_{j=-\infty}^{\infty} E_T[h_{it}k_{i,t-j}]$. Note that $\bar{E}_T[u_{it\alpha}\psi_{is}] = E_T[u_{it\alpha}\psi_{it}]$ if all the regressors are strictly exogenous.[8] Here, σ_i^2 and β_i are components of a higher order expansion for the estimator of the individual effects, that is as $T\to\infty$

$$\begin{aligned} \widehat{\alpha}_i &= \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + o_p(1/T), \ \psi_i = \sum_{t=1}^{T} \psi_{it}/\sqrt{T} \xrightarrow{d} \mathcal{N}(0,\sigma_i^2), && (3.9)\\ \psi_{it} &= -E_T[v_{it\alpha}]^{-1} v_{it}, \quad \sigma_i^2 = \bar{E}_T[\psi_{it}\psi_{is}], && (3.10)\\ \beta_i &= -E_T[v_{it\alpha}]^{-1} \left\{ \bar{E}_T[v_{it\alpha}\psi_{is}] + \frac{1}{2}\sigma_i^2 E_T[v_{it\alpha\alpha}] \right\}, && (3.11) \end{aligned}$$

where again the spectral expectations $\bar{E}_T$ can be replaced by E_T if the regressors are exogenous.

3.4 Bias for Panel Probit Model

The expression for the bias takes a simple form for the static panel probit model with exogenous regressors, which helps explain the results of previous Monte Carlo studies (Greene, 2002; and Hahn and Newey, 2004). In particular, this bias can be expressed as a matrix-weighted average of the true parameter value, where the weighting matrices are positive definite. This implies that probit fixed effects estimators are biased away from zero if the regressor is scalar (as in the studies aforementioned). This property also holds regardless of the dimension of the regressor vector in the absence of heterogeneity, because in this case the weighting matrix is a scalar multiple of the identity matrix. In general the probit fixed effects estimator is upward-biased for θ_0 in the sense of Chamberlain and Leamer (1976) up to order $O_p(T^{-2})$, namely its probability limit θ_T can be expresses as $\theta_T = (\mathcal{I}_p + \mathcal{B}/T)\theta_0 + O_p(T^{-2})$, where $\mathcal{I}_p$ denotes the $p\times p$ identity matrix and $\mathcal{B}$ is a positive definite matrix. Matrix-weighted averages, however, are difficult to interpret and sign except in special cases. These results are stated in the following proposition:

Proposition 1 (Bias for Model Parameters) *Assume that (i) $\{\epsilon_{it}\}_{t=1}^{T}$ is a sequence of random variables such that $\epsilon_{it}|X_i^T, \alpha_i \sim$ i.i.d. $\mathcal{N}(0,1)$ for each i, (ii) $E_T[X_{it}X_{it}']$ exists and is nonsingular for each i, (iii) $\{Z_{it}\}_{t=1}^{T}$ is stationary and strongly mixing with mixing coefficients $a_i(m)$*

[8] The previous expansion provides an alternative explanation for the absence of incidental parameters bias in the panel linear model with exogenous regressors. In this case $v_{it} = Y_{it} - X_{it}'\theta - \alpha_i$, $v_{it\alpha} = -1$, $v_{it\alpha\alpha} = 0$, and $u_{it} = v_{it}X_{it}$. Then, $\beta_i = b = 0$ since $E_T[v_{it\alpha}\psi_{it}] = E_T[u_{it\alpha}\psi_{it}] = 0$ and $E_T[v_{it\alpha\alpha}] = E_T[u_{it\alpha\alpha}] = 0$. Moreover, the bias terms of higher order are also zero because the second order expansions are exact.

such that $\sup_i |a_i(m)| \leq Ca^m$ *for some* a *such that* $0 < a < 1$ *and some* $C > 0$*, (iv)* $\{Z_i^T\}_{i=1}^n$ *are independent, (v)* $\sup_i E_T \left[\|X_{it}\|^{120}\right] < \infty$*, and (vi)* $n = o(T^3)$.[9] *Then,*
1.

$$\mathcal{B} = \frac{1}{2} E_n \left[\mathcal{J}_i\right]^{-1} E_n \left[\sigma_i^2 \mathcal{J}_i\right] \theta_0, \tag{3.12}$$

where

$$\mathcal{J}_i = E_T \left[H_{it}\phi_{it} X_{it} X_{it}'\right] - \sigma_i^2 E_T \left[H_{it}\phi_{it} X_{it}\right] E_T \left[H_{it}\phi_{it} X_{it}'\right], \tag{3.13}$$

$$\sigma_i^2 = E_T \left[H_{it}\phi_{it}\right]^{-1}, \tag{3.14}$$

with $H_{it} = \phi_{it}/[\Phi_{it}(1 - \Phi_{it})]$, $\phi_{it} = \phi(X_{it}'\theta_0 + \alpha_i)$, $\Phi_{it} = \Phi(X_{it}'\theta_0 + \alpha_i)$*, and* ϕ *and* Φ *denote the pdf and cdf of the standard normal distribution.*
2. $E_n \left[\mathcal{J}_i\right]^{-1} \sigma_i^2 \mathcal{J}_i$ *is positive definite for all* i*.*
3. If $\alpha_i = \alpha \ \forall i$*, then*

$$\mathcal{B} = \frac{1}{2} \sigma^2 \theta_0, \tag{3.15}$$

where $\sigma^2 = E_T \left\{\phi(X_{it}'\theta_0 + \alpha)^2 / \left[\Phi(X_{it}'\theta_0 + \alpha)\left(1 - \Phi(X_{it}'\theta_0 + \alpha)\right)\right]\right\}^{-1}$.

Proof. See Appendix C. ■

Condition (i) is the probit modeling assumption; condition (ii) is standard for MLE (Newey and McFadden, 1994), and guarantees identification and asymptotic normality for MLEs of model parameters and individual effects based on time series variation; assumptions (iii), together with the moment condition (v) and (iv), are imposed in order to apply a Law of Large Numbers to the cross sections; and assumptions (v) and (vi) guarantee the existence of, and uniform convergence of remainder terms in, the higher-order expansion of the bias (see, e.g., example 1 in Hahn and Kuersteiner, 2003). Note that the second result follows because σ_i^2 is the asymptotic variance of the estimator of the individual effect α_i, and $\mathcal{J}_i$ corresponds to the contribution of individual i to the inverse of the asymptotic variance of the estimator of the model parameter θ_0. If all these $\mathcal{J}_i$ matrices are diagonal, then $\mathcal{B}$ lies in a ray that starting from the origin passes through θ_0. All the components of the fixed effects estimators are then bias away from zero. Moreover, since $\sigma_i^2 \geq \Phi(0)\left[1 - \Phi(0)\right]/\phi(0)^2 = \pi/2$, the bias $\mathcal{B}$ can be generally bounded from below.

Corollary 1 *Under the conditions of Proposition 1*

$$\|\mathcal{B}\| \geq \frac{\pi}{4} \|\theta_0\|. \tag{3.16}$$

[9] A sequence $\{X_t\}$ is strongly mixing with mixing coefficients $a(m)$ if $\sup_t \sup_{A \in \mathcal{A}_t, B \in \mathcal{B}_{t+m}} |P(A \cap B) - P(A)P(B)| = a(m)$, where $\mathcal{A}_t = \sigma(X_t, X_{t+1}, ...)$ and $\mathcal{B}_t = \sigma(X_t, X_{t+1}, ...)$. $\|\cdot\|$ denotes the Euclidean norm.

When the regressor is scalar or there is no heterogeneity, this lower bound establishes that the first order bias for each index coefficient is at least $\pi/8 \approx 40\%$, $\pi/16 \approx 20\%$ and $\pi/32 \approx 10\%$ for panels with 2, 4 and 8 periods, respectively. In general, these bounds apply to the norm of the coefficient vector. Tighter bounds can also be established for the proportionate bias, $\|\mathcal{B}\|/\|\theta_0\|$, as a function of the true parameter value θ_0. These bounds, however, depend on the joint distribution of regressor and individual effects, and are therefore application specific. Thus, using standard matrix algebra results (see, e.g., Rao, 1973, p. 74), the proportionate bias can be bounded from below and above by the minimum and maximum eigenvalues of the matrix $E_n\left[\mathcal{J}_i\right]^{-1} E_n\left[\sigma_i^2 \mathcal{J}_i\right]/2$, for any value of the parameter vector θ_0.[10]

The third result of the proposition establishes that the bias is proportional to the true parameter value in the absence of heterogeneity. Some intuition for this result can be obtained by analogy to the panel linear model. Specifically, suppose that $Y_{it} = X_{it}'\beta_0 + \alpha_i + \epsilon_{it}$, where $\epsilon_{it} \sim i.i.d.(0, \sigma_\epsilon^2)$. Next, note that in the probit the index coefficients are identified only up to scale, that is $\theta_0 = \beta_0/\sigma_\epsilon$. The probability limit of the fixed effects estimator of this quantity in the linear model, as $n \to \infty$, is

$$\widehat{\theta} = \frac{\widehat{\beta}}{\widehat{\sigma}_\epsilon} \xrightarrow{p} \frac{\beta_0}{\sqrt{1 - 1/T}\sigma_\epsilon} = \left[1 + \frac{1}{2T}\right]\theta_0 + O_p(T^{-2}), \tag{3.17}$$

where the last equality follows from a standard Taylor expansion of $(1-1/T)^{-1/2}$ around $1/T = 0$. Here we can see the parallel with the probit, where $\theta_T = \left[1 + \sigma^2/2T\right]\theta_0 + O(T^{-2})$. Hence, we can think of the bias as coming from the estimation of σ_ϵ, which in the probit case cannot be separated from the estimation of β_0. In other words, the over-fitting due to the fixed effects biases upwards the estimates of the model parameters because the standard deviation is implicitly in the denominator of the model parameter estimated by the probit. The sign of this bias contrasts with Yatchew and Griliches (1985) result about neglected heterogeneity in probit models. They find that estimates that do not account for heterogeneity suffer from attenuation biases if the omitted heterogeneity is normally distributed and independent of the included regressors. I find instead that estimators that account for individual heterogeneity are biased away from zero in the absence of such heterogeneity.

Proportionality implies, in turn, zero bias for fixed effects estimators of ratios of index coefficients. These ratios are often behavioral parameters of interest because they are direct measures of the relative effect of the regressors, and can be interpreted as marginal rates of substitution in economic applications.

[10]Chesher and Jewitt (1987) use a similar argument to bound the bias of the Eicker-White heteroskedasticity consistent covariance matrix estimator.

Corollary 2 *Assume that the conditions of the Proposition 1 hold and* $\alpha_i = \alpha \ \ \forall i$*. Then, for any* $j \neq k \in \{1, ..., p\}$ *and* $\theta = (\theta_1, ..., \theta_p)$

$$\frac{\widehat{\theta}_j}{\widehat{\theta}_k} \xrightarrow{p} \frac{\theta_{0,j}}{\theta_{0,k}} + O_p(T^{-2}). \tag{3.18}$$

In general, the first term of the bias is different for each coefficient depending on the distribution of the individual effects, and the relationship between regressors and individual effects; and shrinks to zero with the inverse of the variance of the underlying distribution of individual effects.

4 Fixed Effects Estimation of Marginal Effects

4.1 Bias Corrections for Marginal Effects

Fixed effects estimators for average marginal effects can be constructed from equations (2.7), (2.8), and (2.9) by replacing population moments by sample analogs and using fixed effects estimators of the individual effects, that is

$$\widehat{\mu}(x_1, \theta) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} m\left((x_1, X_{2it}), \theta, \widehat{\alpha}_i(\theta)\right), \tag{4.1}$$

$$\widehat{\mu}(\theta) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{m}\left(X_{it}, \theta, \widehat{\alpha}_i(\theta)\right), \tag{4.2}$$

and

$$\widehat{\mu}_1(\theta) = \frac{1}{N_1} \sum_{i=1}^{n} \sum_{t=1}^{T} m\left((0, X_{2it}), \theta, \widehat{\alpha}_i(\theta)\right) \mathbf{1}\left\{X_{1it} = 1\right\}, \tag{4.3}$$

where $N_1 = \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}\{X_{1it} = 1\}$.

These fixed effects estimators, however, suffer also from the incidental parameters problem even if they are evaluated at the true value of the index coefficients θ_0. The source of this bias is the dependence on the estimators of the individual effects $\widehat{\alpha}_i(\theta)$. As for model parameters, the slow rate of convergence of these estimators introduces bias in the asymptotic distribution of the estimators of the marginal effects. In particular, from the asymptotic expansion for $\widehat{\alpha}_i(\theta_0)$ in (3.9) we have, as $n \to \infty$ and $T \to \infty$,

$$\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{m}\left(X_{it}, \theta_0, \widehat{\alpha}_i(\theta_0)\right) \xrightarrow{p} E_n\left\{E_T\left[\tilde{m}\left(X_{it}, \theta_0, \alpha_i\right)\right] + \frac{1}{T} E_T\left[\Delta_i\right] + o_p(T^{-1})\right\}, \tag{4.4}$$

where

$$\Delta_i = E_T\left[\tilde{m}_\alpha\left(X_{it}, \theta_0, \alpha_i\right)\right] \beta_i + \bar{E}_T\left[\tilde{m}_\alpha\left(X_{it}, \theta_0, \alpha_i\right) \psi_{is}\right] + \frac{1}{2} E_T\left[\tilde{m}_{\alpha\alpha}\left(X_{it}, \theta_0, \alpha_i\right)\right] \sigma_i^2, \tag{4.5}$$

and subscripts on $\tilde{m}$ denote partial derivatives.[11] Note that if all the regressors are exogenous the term that includes the influence function ψ_{is} drops out and the previous equation reduces to the expression given in HN. Appendix B describes how to construct bias corrected estimates of average marginal effects from (4.4) and (4.5) by replacing population moments for sample analogs and index coefficients for their bias corrected estimates.

4.2 Panel Probit: Small Bias Property

Bias-corrected estimators of marginal effects are consistent up to order $O_p\left(T^{-2}\right)$ and have asymptotic distributions centered at the true parameter value if $T/n^{1/3} \to \infty$. This can be shown using a large T-expansion of the estimator, just as for model parameters. The question addressed here, however, is whether these corrections are indeed needed. In other words, how important is the bias that the corrections aim to remove? This question is motivated by Monte Carlo evidence in Hahn and Newey (2004), which shows negligible biases for estimators of marginal effects evaluated at uncorrected fixed effects estimators of the index coefficients. The following proposition gives the analytical expression for the bias of probit fixed effects estimators of marginal effects in a static model with exogenous regressors.

Proposition 2 (Bias for Marginal Effects) *Let* $\widehat{\mu}(\widehat{\theta}) = \widehat{\theta} \ \sum_{i=1}^{n}\sum_{t=1}^{T}\phi\left(X_{it}'\widehat{\theta} + \widehat{\alpha}_i(\widehat{\theta})\right)/nT$ *and* $\mu = \theta_0 \ E_n\left\{E_T\left[\phi\left(X_{it}'\theta_0 + \alpha_i\right)\right]\right\}$, *then under the conditions of Proposition 1, as* $n, T \to \infty$

$$\widehat{\mu}(\widehat{\theta}) \xrightarrow{p} \mu + \frac{1}{T}\mathcal{B}_\mu + O(T^{-2}), \tag{4.6}$$

where

$$\mathcal{B}_\mu = \frac{1}{2}E_n\left\{E_T\left[\phi_{it}\left(\xi_{it}\theta_0\left(X_{it} - \sigma_i^2 E_T\left[H_{it}\phi_{it}X_{it}\right]\right)' - \mathcal{I}_p\right)\right]\left(\sigma_i^2\mathcal{I}_p - E_n\left[\mathcal{J}_i\right]^{-1}E_n\left[\sigma_i^2\mathcal{J}_i\right]\right)\right\}\theta_0, \tag{4.7}$$

$$\xi_{it} = X_{it}'\theta_0 + \alpha_i, \quad \sigma_i^2 = E_T\left[H_{it}\phi_{it}\right]^{-1}, \tag{4.8}$$

$$\mathcal{J}_i = -\left\{E_T\left[H_{it}\phi_{it}X_{it}X_{it}'\right] - \sigma_i^2 E_T\left[H_{it}\phi_{it}X_{it}\right]E_T\left[H_{it}\phi_{it}X_{it}'\right]\right\}, \tag{4.9}$$

with $H_{it} = \phi_{it}/[\Phi_{it}(1-\Phi_{it})]$, $\phi_{it} = \phi(\xi_{it})$, $\Phi_{it} = \Phi(\xi_{it})$, ϕ *and* Φ *denote the pdf and cdf of the standard normal distribution, and* $\mathcal{I}_p$ *denotes a* $p \times p$ *identity matrix.*

Proof. See Appendix C. ■

When X_{it} is scalar all the formulas can be expressed as functions uniquely of the index ξ_{it} as

$$\mathcal{B}_\mu = E_n\left[\mathcal{B}_{\mu_i}\right] = E_n\left[\delta_i\pi_i\right]\theta_0/2, \tag{4.10}$$

[11] Similar expansions can be obtained for the measures of average marginal effects based on $m(x_{it}, \theta, \alpha_i)$.

with $\delta_i = E_T\left\{\phi_{it}\left[\xi_{it}\left(\xi_{it} - E_T\left[H_{it}\phi_{it}\right]^{-1} E_T\left[H_{it}\phi_{it}\xi_{it}\right]\right) - 1\right]\right\}$, $\pi_i = \sigma_i^2 - E_n\left[\mathcal{J}_i\right]^{-1} E_n\left[\sigma_i^2 \mathcal{J}_i\right]$, and $\mathcal{J}_i = -\left\{E_T\left[H_{it}\phi_{it}\xi_{it}^2\right] - \sigma_i^2 E_T\left[H_{it}\phi_{it}\xi_{it}\right]^2\right\}/\theta_0^2$. Here, we can see that the bias is an even function of the index, ξ_{it}, since all the terms are centered around weighted means that are symmetric around the origin.[12] For the term π_i, each weight $\mathcal{J}_i$ is the residual variance of a weighted regression of the index ξ_{it} on a constant with weights $\sqrt{H_{it}\phi_{it}}$, whereas σ_i^2 is the sum of squared weights. Hence we have that $\pi_i \approx \sigma_i^2 - E_n[\sigma_i^2]$, which has is an even function on α_i around the origin with zero mean. The additional weights δ_i are also even functions of α_i around the origin, preserving the approximate zero mean property. This is illustrated in Figure 3, which plots the components of the bias for independent normally distributed regressor and individual effect. The terms δ_i and π_i have U shapes and take the same sign for $\alpha_i = 0$ (when the mean of the regressors is absorbed in the individual effects); the term ϕ_{it} in δ_i acts to reduce the weights in the tails, where the other components are large. In this case the bias, as a function of α, is positive at zero and takes negative values as we move away from the origin. Then, positive and negative values compensate each other when they are integrated using the distribution of the individual effect to obtain $\mathcal{B}_\mu$.

The following example illustrates the argument of the proof (in Appendix C) and shows a case where this bias is exactly zero.

Example 1 (Panel probit without heterogeneity) *Consider a probit model where the individual effect is the same for all the individuals, that is $\alpha_i = \alpha \;\; \forall i$. In this case, as $n, T \to \infty$*

$$\widehat{\mu} \xrightarrow{p} \mu + O_p(T^{-2}). \tag{4.11}$$

First, note that as $n, T \to \infty$

$$\widehat{\mu} \xrightarrow{p} E_n\left\{\theta_T E_T\left[\phi\left(X_{it}'\theta_T + \widehat{\alpha}_i(\theta_T)\right)\right]\right\}. \tag{4.12}$$

Next, in the absence of heterogeneity (see Proposition 1, and proof of Proposition 2 in Appendix C)

$$\theta_T = \theta_0 + \frac{1}{2T}\sigma^2\theta_0 + O\left(T^{-2}\right),\ \widehat{\alpha}_i(\theta_T) = \alpha + \psi_i/\sqrt{T} + \alpha\sigma^2/2T + R_i/T^{3/2}, \tag{4.13}$$

where under the conditions of Proposition 2

$$\psi_i \overset{a}{\sim} N(0, \sigma^2),\ \sigma^2 = E_T\left[H_{it}\phi_{it}\right]^{-1}, \quad E_n\left[R_i\right] = O(T^{-1/2}). \tag{4.14}$$

Combining these results, the limit of the index, $\widehat{\xi}_{it} = X_{it}'\theta_T + \widehat{\alpha}_i(\theta_T)$, has the following expansion

$$\widehat{\xi}_{it} = \left(1 + \sigma^2/2T\right)\xi_{it} + \psi_i/\sqrt{T} + R_i/T^{3/2} + O(T^{-2}), \quad \xi_{it} = X_{it}'\theta_0 + \alpha. \tag{4.15}$$

[12] A function $f : \mathbb{R}^n \to \mathbb{R}$ is *even* if $f(-x) = f(x)\ \forall x \in \mathbb{R}^n$.

Finally, replacing this expression in (4.12), using the convolution properties of the normal distribution, see Lemma 1 in Appendix C, and assuming that orders in probability correspond to orders in expectation,

$$\begin{aligned}\widehat{\mu} \quad \xrightarrow{p} \quad & E_n\left\{\theta_T \; E\left[\phi(\widehat{\xi}_{it})|X_i^T,\alpha\right]\right\} = E_n\left[\theta_T \; \int \phi\left(\left(1+\frac{\sigma^2}{2T}\right)\xi_{it}+v\right)\frac{\sqrt{T}}{\sigma}\;\phi\left(\frac{\sqrt{T}v}{\sigma}\right)dv\right] \\ + \quad & O(T^{-2}) = E_n\left[\frac{\left(1+\sigma^2/2T\right)}{\sqrt{1+\sigma^2/T}}\;\theta_0\;\phi\left(\frac{\left(1+\sigma^2/2T\right)\xi_{it}}{\sqrt{1+\sigma^2/T}}\right)\right] + O(T^{-2}) = \mu + O(T^{-2}),\end{aligned} \tag{4.16}$$

since by a standard Taylor expansion

$$\left(1+\sigma^2/T\right)^{-1/2}\left(1+\sigma^2/2T\right) \quad = \quad \left(1-\frac{\sigma^2}{2T}+O(T^{-2})\right)\left(1+\frac{\sigma^2}{2T}\right) = 1 + O(T^{-2}). \tag{4.17}$$

In other words, the standard deviation of the random part of the limit index exactly compensates for the first term of the bias in the conditional expectation of the nonlinear function $\phi(\cdot)$.

This example shows that, as in linear models, the inclusion of irrelevant variables, while reducing efficiency, does not affect the consistency of the probit estimates of marginal effects. Moreover, this example complements Wooldridge's (2002, Ch. 15.7.1) result about neglected heterogeneity in panel probit models. Wooldridge shows that estimates of average effects that do not account for unobserved heterogeneity are consistent, if the omitted heterogeneity is normally distributed and independent of the included regressors. Here, on the other hand, I find that estimates of marginal effects that account for heterogeneity are consistent in the absence of such heterogeneity.

In general, the bias depends upon the degree of heterogeneity and the joint distribution of regressors and individual effects. Table 1 reports Monte Carlo estimates of the biases of fixed effects estimators of index coefficients and marginal effects, and numerical estimates of their first order approximation based on the expressions in Propositions 1 and 2. All the entries of the table are expressed in percent with respect to the true value of the corresponding parameter. The examples correspond to an 8-period model with one regressor, and the model parameter θ_0 equal to 1. I consider several distributions for the regressors and individual effects normalized to have zero mean and unit variance, except for the Nerlove process and $C = 0$ that denotes a degenerate distribution at zero.[13] The bias for index coefficients lies between the 10% lower bound and 20%,

[13]The distribution $C = 0$ corresponds to the case of absence of heterogeneity, i.e., $\alpha_i = \alpha \; \forall i$.

whereas for marginal effects is never higher than 2%. The first order approximation of the bias captures more than 70% of the bias for index coefficients even in this case where T is moderate (8 time periods). For the marginal effects, the first order approximation is relatively less accurate, but in this case the bias is very small.

The small bias property for fixed effects estimators of marginal effects does carry over to dynamic models or more generally to models with predetermined regressors. The reason is that averaging across individuals does not fully remove the additional bias components due to the dynamic feedbacks. To understand this result, we can look at the survivor probabilities at zero in a dynamic Gaussian linear model. Specifically, suppose that $Y_{it} = \theta_0 Y_{i,t-1} + \alpha_i + \epsilon_{it}$, where $\epsilon_{it}|Y_{i,t-1}, ..., Y_{i,0}, \ \alpha_i \sim N(0, \sigma^2)$, $Y_{i0}|\alpha_i \sim \mathcal{N}(\alpha_i/(1-\theta_0), \ \sigma^2/(1-\theta_0^2))$, and $0 \leq \theta_0 < 1$. The survivor probability evaluated at $Y_{i,t-1} = r$ and its fixed effects estimator are

$$S = E_n\left\{\Phi\left(\frac{\theta_0 \ r + \alpha_i}{\sigma}\right)\right\}, \quad \widehat{S} = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{\widehat{\theta} \ r + \widehat{\alpha}_i(\widehat{\theta})}{\widehat{\sigma}}\right), \tag{4.18}$$

where $\widehat{\theta}$ and $\widehat{\sigma}^2$ are the fixed effects MLEs of θ_0 and σ^2. It can be shown that $\widehat{\theta}$ converges to $\theta_T = \theta_0 - (1+\theta_0)/T + O(T^{-2})$, and $\widehat{\sigma}^2$ converges to $\sigma_T^2 = \sigma^2 - \sigma^2/T + O(T^{-2})$, as $n, T \to \infty$ (Nickell, 1981). For the estimator of the individual effects, a large-T expansion gives

$$\widehat{\alpha}_i(\theta_T) = \alpha_i + v_i - (\theta_T - \theta_0)\frac{\alpha_i}{1-\theta_0} + o_p(1/T), \quad v_i \sim \mathcal{N}(0, \sigma^2/T). \tag{4.19}$$

Then, as $n, T \to \infty$

$$\begin{aligned}
\widehat{S} \ &\xrightarrow{p} \ E_n\left\{E\left[\Phi\left(\frac{\theta_T \ r + \widehat{\alpha}_i(\theta_T)}{\sigma_T}\right)\right]\right\} \\
&= \ E_n\left\{E\left[\Phi\left(\frac{\theta_0 r + \alpha_i + v_i + (\theta_T - \theta_0)\left(r - \frac{\alpha_i}{1-\theta_0}\right) + o_p(T^{-1})}{\sigma_T}\right)\right]\right\} \\
&= \ E_n\left\{\Phi\left(\frac{\theta_0 r + \alpha_i - \frac{1}{T}(1+\theta_0)\left(r - \frac{\alpha_i}{1-\theta_0}\right) + o(T^{-1})}{\sigma}\right)\right\},
\end{aligned} \tag{4.20}$$

by the convolution properties of the normal distribution, see Lemma 1 in Appendix C.

In expression (4.20) we can see that averaging across individuals eliminates the bias of $\widehat{\sigma}^2$, but does not affect the bias of $\widehat{\theta}$. The sign of the bias of $\widehat{S}$ generally depends on the distribution of the individual effects. When there is no heterogeneity ($\alpha_i = \alpha \ \forall i$), for example, $\widehat{S}$ underestimates (overestimates) the underlying survivor probability when evaluated at values above (below) the unconditional mean of the response, $\alpha/(1-\theta_0)$. This means that if the marginal effects are computed as differences in survivor probabilities evaluated at two different values (see, e.g., expression (2.7) for discrete regressors), fixed effects estimates of marginal effects would be biased downward if the values chosen are $x_1 = 0$ and $x_1 + 1 = 1$. For exogenous variables, Monte Carlo results suggest that the bias problem is less severe (see Section 5). Intuitively, it seems that the part of the bias due to the dynamic feedbacks is less important for this type of regressors.

5 Monte Carlo Experiments

This section reports evidence on the finite sample behavior of fixed effects estimators of model parameters and marginal effects for static and dynamic models. In particular, I analyze the finite sample properties of uncorrected and bias-corrected fixed effects estimators in terms of bias and inference accuracy of the asymptotic distribution. The small bias property for marginal effects is illustrated for several lengths of the panel.[14] All the results presented are based on 1000 replications, and the designs are as in Heckman (1981), Greene (2002), and HN for the static probit model, and as in Honoré and Kyriazidou (2000), Hahn and Kuersteiner (2003), and Carro (2006) for the dynamic logit model.

I introduce new large-T bias corrected estimators for panel logit and probit models that exhibit good finite sample properties relative to the existing alternatives to estimate these models. These corrections are based on using more intensively the parametric structure of the problem and replace observed quantities for expected quantities in the estimation of the bias. This approach is similar to the use of the conditional information matrix in the estimation of asymptotic variances in maximum likelihood problems, instead of other alternatives, such as the sample average of the outer product of the scores or the sample average of the negative Hessian (Porter, 2002). The results below show that this refinement improve the finite sample performance of the correction. The expressions for the bias for binary choice models and the corresponding bias corrections are described in the Appendices A and B.

[14]In results not reported, I check the robustness of the estimators to small deviations from correct specification. Thus, the performance of probit and logit estimators is evaluated when the error term is logistic and normal, respectively. See Fernandez-Val (2005).

5.1 Static Probit Model

The model design is

$$
\begin{aligned}
Y_{it} &= \mathbf{1}\left\{X_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\right\}, \ \epsilon_{it} \sim \mathcal{N}(0,1), \ \alpha_i \sim \mathcal{N}(0,1), && (5.1)\\
X_{it} &= t/10 + X_{i,t-1}/2 + u_{it}, \quad X_{i0} = u_{i0}, \quad u_{it} \sim \mathcal{U}(-1/2, 1/2), && (5.2)\\
n &= 100, \ T = 4, \ 8, \ 12; \ \theta_0 = 1, && (5.3)
\end{aligned}
$$

where $\mathcal{N}$ and $\mathcal{U}$ denote normal and uniform distribution, respectively. Throughout the tables reported, SD is the standard deviation of the estimator; $\widehat{p}; \#$ denotes a rejection frequency with # specifying the nominal value; SE/SD is the ratio of the average standard error to standard deviation; and MAE denotes median absolute error.[15] $BC1$ and $BC2$ correspond to the one-step analytical bias-corrected estimators of HN based on maximum likelihood setting and general estimating equations, respectively. JK is the bias correction based on the leave-one-period-out Jackknife-type estimator, see HN. $BC3$ is the one-step bias-corrected estimator proposed here. Iterated bias-corrected and score-corrected estimators are not considered because they are much more cumbersome computationally.[16]

Table 2 gives the Monte Carlo results for the estimators of θ_0 when ϵ_{it} is normally distributed. The results here are similar to previous studies (Greene, 2002; and HN) and show that the probit MLE is severely biased, even when $T = 12$, and has important distortions in rejection probabilities. $BC3$ reduces substantially the bias, and improves in terms of bias and rejection probabilities over HN's analytical and jackknife bias-corrected estimators for small sample sizes.[17] The corrections reduce dispersion, which can be explained by the proportionality result for the bias in Proposition 2.

Table 3 reports the ratio of estimators to the truth for marginal effects. Here, I include also two estimators of the average effect based on linear probability models. $LPM - FS$ is the standard linear probability model that uses all the observations; LPM is an adjusted version that calculates the slope from individuals that change status during the sample, i.e., excluding individuals with $Y_{it} = 1 \ \forall t$ or $Y_{it} = 0 \ \forall t$, and assigns zero effect to the rest. The results show small biases in uncorrected fixed effects estimators of marginal effects. Rejection frequencies are higher than their nominal levels, due to underestimation of dispersion. As in cross-section

[15] I use median absolute error instead of root mean squared error as an overall measure of goodness of fit because it is less sensitive to outliers.

[16] HN find that iterating the bias correction does not matter much in this example.

[17] Note that the maximal simulation standard errors for the mean estimates are 1%, .5%, and .3% for 4, 8, and 12 time periods, respectively

models (Angrist, 2001, and Hahn, 2001), both linear models work fairly well in estimating the average effect.[18]

5.2 Dynamic Logit Model

The model design is

$$Y_{i0} = \mathbf{1}\{\theta_{X,0}X_{i0} + \alpha_i - \epsilon_{i0} \geq 0\}, \tag{5.4}$$

$$Y_{it} = \mathbf{1}\{\theta_{Y,0}Y_{i,t-1} + \theta_{X,0}X_{it} + \alpha_i - \epsilon_{it} \geq 0\}, \quad t = 1, ..., T-1, \tag{5.5}$$

$$\epsilon_{it} \sim \mathcal{L}(0, \pi^2/3), \quad X_{it} \sim \mathcal{N}\left(0, \pi^2/3\right), \tag{5.6}$$

$$n = 250;\ T = 8,\ 12,\ 16;\ \theta_{Y,0} = .5;\ \theta_{X,0} = 1, \tag{5.7}$$

where $\mathcal{L}$ denotes the standardized logistic distribution. Here, the individual effects are correlated with the regressor. In particular, to facilitate the comparison with other studies, I follow Honoré and Kyriazidou (2000) and generate $\alpha_i = \sum_{t=0}^{3} X_{it}/4$. The measures reported are the same as for the static case. $BC1$ denotes the bias-corrected estimator of Hahn and Kuersteiner (2003); HK is the dynamic version of the conditional logit of Honoré and Kyriazidou (2000), which is fixed-T consistent; MML is the Modified MLE for dynamic models of Carro (2006); and $BC3$ is the bias-corrected estimator that uses expected quantities in the estimation of the bias formulas.[19] For the number of lags, I choose a bandwidth parameter $J = 1$, as in Hahn and Kuersteiner (2003).

Tables 4 and 5 present the Monte Carlo results for the structural parameters $\theta_{Y,0}$ and $\theta_{X,0}$. Overall, all the bias-corrected estimators have smaller finite sample bias and better inference properties than the uncorrected MLEs. Large-T-consistent estimators have median absolute error comparable to HK for $T = 8$.[20] Among them, $BC3$ and MML are slightly superior to $BC1$, but there is no clear ranking between them. Note, however, that $BC3$ allows for regressors to be predetermined, and is computationally more attractive than MML as it does not require modification of the probit first order conditions.

Tables 6 and 7 report the Monte Carlo results for ratios of the estimator to the truth for average effects for the lagged dependent variable and exogenous regressor, respectively. These

[18]Stoker (1986) derives the expression for the OLS estimand in index models (e.g., probit and logit). This estimand corresponds to the average effect under normality of regressors and individual effects.

[19]HK and MML results are extracted from the tables reported in their articles and therefore some of the measures are not available. HK results are based on a bandwidth parameter equal to 8.

[20]An aspect not explored here is that the performance of HK estimator deteriorates with the number of exogenous variables. Thus, Hahn and Kuersteiner (2003) find that their large-T-consistent estimator out-performs HK for $T = 8$ when the model includes two exogenous variables.

effects are calculated using expression (4.1) with $x_1 = 0$ for the lagged dependent variable, and expression (4.2) for the exogenous regressor. Here, I present results for MLE, $BC1$, $BC3$, linear probability models (LPM and $LPM - FS$), and bias-corrected linear models ($BC - LPM$ and $BC - LPM - FS$) constructed using Nickell's (1981) bias formulas. As in the example of the linear model in Section 4, uncorrected estimates of the effects of the lagged dependent variable are biased downward. Uncorrected estimates of the effect for the exogenous variable, however, have small biases. Large-T corrections are effective in reducing bias and fixing rejection probabilities for both linear and nonlinear estimators.

6 Empirical Illustration: Female Labor Force Participation

The relationship between fertility and female labor force participation is of longstanding interest in labor economics and demography. For a recent discussion and references to the literature, see Angrist and Evans (1998). Research on the causal effect of fertility on labor force participation is complicated because both variables are jointly determined. In other words, there exist multiple unobserved factors (to the econometrician) that affect both decisions. Here, I adopt an empirical strategy that aims to solve this omitted variables problem by controlling for unobserved individual time-invariant characteristics using panel data. Other studies that follow a similar approach include Heckman and MaCurdy (1980), Heckman and MaCurdy (1982), Hyslop (1999), Chay and Hyslop (2000), Carrasco (2001), and Carro (2006).

The empirical specification I use is similar to Hyslop (1999). In particular, I estimate the following equation

$$P_{it} = \mathbf{1}\left\{\delta_t + P_{i,t-1}\theta_P + X_{it}'\theta_X + \alpha_i - \epsilon_{it} \geq 0\right\}, \tag{6.1}$$

where P_{it} is the labor force participation indicator; δ_t is a period-specific intercept; $P_{i,t-1}$ is the participation indicator of the previous period; and X_{it} is a vector of time-variant covariates that includes three fertility variables - the numbers of children aged 0-2, 3-5, and 6-17 -, log of husband's earnings, and a quadratic function of age.[21]

The sample is selected from waves 13 to 22 of the Panel Study of Income Dynamics (PSID) and contains information of the ten calendar years 1979-1988. Only women aged 18-60 in 1985, continuously married, and whose husband is in the labor force in each of the sample periods are included in the sample. The final sample consists of 1,461 women, 664 of whom change labor

[21] Hyslop (1999) specification includes also the lag of the number of 0 to 2 year-old children as additional regressor. This regressor, however, is statistically nonsignificant at the 10% level.

force participation status during the sample period. The first year is excluded to use it as initial condition for the dynamic model.

Descriptive statistics for the sample are shown in Table 8. Twenty-one percent of the sample is black, and the average age in 1985 was 37. Roughly 72% of women participate in the labor force at some period, the average schooling is 12 years, and the average numbers of children are .2, .3 and 1.1 for the three categories 0-2 year-old, 3-5 year-old, and 6-17 year-old children, respectively.[22] Women that change participation status during the sample, in addition to be younger, less likely to be black, and less educated, have more dependent children and their husband's earnings are slightly higher than average. Interestingly, women who never participate do not have more children than women who are employed each year, though this can be explained in part by the non-participants being older. All the covariates included in the empirical specification display time variation over the period considered.

Table 9 reports fixed effects estimates of index coefficients and marginal effects obtained from a static specification, that is, excluding the lag of participation in equation (6.1). Estimators are labeled as in the Monte Carlo example, with C denoting Andersen (1973) conditional logit estimator and $BC3_p$ the bias corrected estimator when the regressors are treated as predetermined.[23] The results show that uncorrected estimates of index coefficients are about 15 percent larger than their bias-corrected counterparts; whereas the corresponding differences for marginal effects are less than 2 percent, and insignificant relative to standard errors. It is also remarkable that all the corrections considered give very similar estimates for both index coefficients and marginal effects (for example, bias-corrected logit estimates are the same as conditional logit estimates, up to two decimal points).[24] The adjusted linear probability model gives estimates of the marginal effects closer to logit and probit than the standard linear model. According to the static model estimates, an additional child aged less than 2 reduces the probability of participation by 9 percent, while each child aged 3-5 and 6-17 reduces the probability of participation

[22] Years of schooling is imputed from the following categorical scheme: 1 = '0-5 grades' (2.5 years); 2 = '6-8 grades' (7 years); 3 = '9-11 grades' (10 years); 4 = '12 grades' (12 years); 5 = '12 grades plus nonacademic training' (13 years); 6 = 'some college' (14 years); 7 = 'college degree' (15 years); 7 = 'college degree, not advanced' (16 years); 8 = 'college and advanced degree' (18 years). See also Hyslop (1999).

[23] Technically the time dummies do not satisfy the regularity conditions for the validity of the large-T bias corrections because they are also incidental parameters under large-T asymptotics. In results not reported, I find that excluding the time dummies does not have any significant effect on the estimates. These results are available from the author upon request. Extending the large-T bias corrections to the presence of time dummies is an open question and is the object of current research by the author.

[24] Logit index coefficients are multiplied by $\sqrt{3}/\pi$ to have the same scale as probit index coefficients.

by 5 percent and 2 percent, respectively. Allowing for feedback from the endogenous response to the regressors has very little effect on the estimates.

In the presence of positive state dependence, estimates from a static model overstate the effect of fertility because additional children reduce the probability of participation and participation is positively serially correlated. This can be seen in Table 10, which reports fixed effects estimates of index coefficients and marginal effects using a dynamic specification. Here, as in the Monte Carlo example, uncorrected estimates of the index coefficient and marginal effect of the lagged dependent variable are significantly smaller (relative to standard errors) than their bias-corrected counterparts for both linear and nonlinear models. Bias-corrected probit gives estimates of index coefficients very similar to probit Modified Maximum Likelihood.[25] The adjusted linear probability model, again, gives estimates of the average effects closer to logit and probit than the standard linear model. Each child aged 0-2 and 3-5 reduces the probability of participation by 6 percent and 3 percent, respectively; while an additional child aged more than 6 years does not have a significant effect on the probability of participation (at the 5 percent level). Finally, a one percent increase in the income earned by the husband reduces a woman's probability of participation by about 0.03%. This elasticity is not sensitive to the omission of dynamics or to the bias corrections.

7 Summary and conclusions

This paper derives analytical expressions for the bias of fixed effects estimators of index coefficients and marginal effects in probit models. The expression for the index coefficients shows that the bias is proportional to the true value of the parameter and can be bounded from below. Moreover, fixed effects estimators of ratios of coefficients and marginal effects do not suffer from the incidental parameters problem in the absence of heterogeneity, and generally have much smaller biases than fixed effects estimators of the index coefficients. These results are illustrated with Monte Carlo examples and an empirical application that analyzes female labor force participation using data from the PSID.

It would be useful to know if the small bias property of fixed effects estimators of average effects generalizes to other statistics of the distribution of effects in the population, like median effects or other quantile effects. However, such analysis is expected to be more complicated because these statistics are non-smooth functions of the data and therefore the standard expansions cannot be directly used. I leave this analysis for future research.

[25] Modified Maximum Likelihood estimates are taken from Carro (2006)

Appendix

A Bias Formulas for Binary Choice Models

The conditional log-likelihood and the scores for observation i at time t are

$$l_{it}(\theta,\alpha_i) = Y_{it}\log F_{it}(\theta,\alpha_i) + (1-Y_{it})\log(1-F_{it}(\theta,\alpha_i)), \tag{A.1}$$

$$v_{it}(\theta,\alpha_i) = H_{it}(\theta,\alpha_i)\,(Y_{it}-F_{it}(\theta,\alpha_i)), \quad u_{it}(\theta,\alpha_i) = v_{it}(\theta,\alpha_i)X_{it}, \tag{A.2}$$

where $F_{it}(\theta,\alpha_i)$ denotes $F_\epsilon(X_{it}'\theta+\alpha_i|X_i^t,\alpha_i)$, $H_{it}(\theta,\alpha_i) = f_{it}(\theta,\alpha_i)/\left[F_{it}(\theta,\alpha_i)\,(1-F_{it}(\theta,\alpha_i))\right]$, $f_{it}(\theta,\alpha_i)$ denotes $f_\epsilon(X_{it}'\theta+\alpha_i|X_i^t,\alpha_i)$, and f_ϵ is the pdf associated to F_ϵ.

Next, since by the Law of Iterated Expectations $E_T\left[h(Z_{it})\right] = E_T\left[E\left[h(Z_{it})|X_i^t,\alpha\right]\right]$ for any function $h(Z_{it})$, taking conditional expectations of the expressions for the components of the bias in Section 3 yields

$$\sigma_i^2 = E_T\left[H_{it}f_{it}\right]^{-1}, \quad \beta_i = -\sigma_i^4 E_T\left[H_{it}g_{it}\right]/2 - \sigma_i^2\tilde{E}_T[H_{it}f_{it}\psi_{is}], \tag{A.3}$$

$$b = -E_n\left\{E_T\left[H_{it}f_{it}X_{it}\right]\beta_i + E_T\left[H_{it}g_{it}X_{it}\right]\sigma_i^2/2 - \tilde{E}_T\left[H_{it}f_{it}X_{it}\psi_{is}\right]\right\}, \tag{A.4}$$

$$\mathcal{J} = -E_n\left\{E_T\left[H_{it}f_{it}X_{it}X_{it}'\right] - \sigma_i^2 E_T\left[H_{it}f_{it}X_{it}\right]E_T\left[H_{it}f_{it}X_{it}'\right]\right\}, \tag{A.5}$$

where $\tilde{E}_T\left[h_{it}k_{is}\right] := \sum_{j=1}^{\infty} E_T\left[h_{it}k_{i,t-j}\right]$, $g_{it}(\theta,\alpha_i)$ denotes $g_\epsilon(X_{it}'\theta+\alpha_i|X_i^t,\alpha_i)$, g_ϵ is the derivative of f_ϵ, and all the expressions are evaluated at the true parameter value (θ_0,α_i). If the regressors are exogenous, the terms involving $\tilde{E}_T$ drop from the previous expressions.

B Bias Corrections for Binary Choice Models

B.1 Bias Corrections for Model Paramenters

The expressions in Appendix A can be used to construct bias-corrected estimators for model parameters in panel binary choice models. Let

$$\widehat{F}_{it}(\theta) := F_{it}(\theta,\widehat{\alpha}_i(\theta)), \quad \widehat{f}_{it}(\theta) := f_{it}(\theta,\widehat{\alpha}_i(\theta)), \quad \widehat{g}_{it}(\theta) := g_{it}(\theta,\widehat{\alpha}_i(\theta)), \quad \widehat{H}_{it}(\theta) := H_{it}(\theta,\widehat{\alpha}_i(\theta)). \tag{B.1}$$

Also, define

$$\widehat{\sigma}_i^2(\theta) := \widehat{E}_T\left[\widehat{H}_{it}(\theta)\,\widehat{f}_{it}(\theta)\right]^{-1}, \quad \widehat{\psi}_{it}(\theta) := \widehat{\sigma}_i^2(\theta)\widehat{H}_{it}(\theta)\left[Y_{it}-\widehat{F}_{it}(\theta)\right], \tag{B.2}$$

where $\widehat{E}_T[h_{it}] := \sum_{t=1}^{T} h_{it}/T$. Here, $\widehat{\sigma}_i^2(\theta)$ and $\widehat{\psi}_{it}(\theta)$ are estimators of the asymptotic variance and influence function, respectively, obtained from a expansion of $\widehat{\alpha}_i(\theta)$ as T grows, after taking conditional expectations given X_i^t and α_i, see (3.9) and the expressions in Appendix A. Let

$$\widehat{\beta}_i(\theta) := -\widehat{\sigma}_i^4(\theta)\widehat{E}_T\left[\widehat{H}_{it}(\theta)\,\widehat{g}_{it}(\theta)\right]/2 - \widehat{\sigma}_i^2(\theta)\widehat{\tilde{E}}_{T,J}[\widehat{H}_{it}(\theta)\,\widehat{f}_{it}(\theta)\,\widehat{\psi}_{is}(\theta)], \tag{B.3}$$

$$\widehat{\mathcal{J}}(\theta) := -\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{E}_T\left[\widehat{H}_{it}(\theta)\widehat{f}_{it}(\theta)X_{it}X_{it}'\right] - \widehat{\sigma}_i^2(\theta)\widehat{E}_T\left[\widehat{H}_{it}(\theta)\widehat{f}_{it}(\theta)X_{it}\right]\widehat{E}_T\left[\widehat{H}_{it}(\theta)\,\widehat{f}_{it}(\theta)X_{it}'\right]\right\}, \tag{B.4}$$

where $\widehat{\widetilde{E}}_{T,J}[h_{it}k_{is}] := \sum_{j=1}^{J}\sum_{t=j+1}^{T} h_{it}k_{i,t-j}/(T-j)$ and J is a bandwidth parameter that needs to be chosen such that $J/T^{1/2} \to 0$ as $T \to \infty$, see Hahn and Kuersteiner (2003). Here, $\widehat{\beta}_i(\theta)$ is an estimator of the higher-order asymptotic bias of $\widehat{\alpha}_i(\theta)$ from a stochastic expansion as T grows, and $\widehat{\mathcal{J}}(\theta)$ is an estimator of the Jacobian of the estimating equation for θ. Then, the estimator of $\mathcal{B}$ is

$$\widehat{\mathcal{B}}(\theta) = -\widehat{\mathcal{J}}(\theta)^{-1}\widehat{b}(\theta), \tag{B.5}$$

where

$$\widehat{b}(\theta) := -\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{E}_T\left[\widehat{H}_{it}(\theta)\widehat{f}_{it}(\theta)X_{it}\right]\widehat{\beta}_i(\theta) + \widehat{E}_T\left[\widehat{H}_{it}(\theta)\widehat{g}_{it}(\theta)X_{it}\right]\widehat{\sigma}_i^2(\theta)/2 - \widehat{\widetilde{E}}_{T,J}\left[\widehat{H}_{it}(\theta)\widehat{f}_{it}(\theta)X_{it}\widehat{\psi}_{is}(\theta)\right]\right\}, \tag{B.6}$$

is an estimator of the bias of the estimating equation for θ, and J is again a bandwidth parameter chosen such that $J/T^{1/2} \to 0$ as $T \to \infty$.

One step bias corrected estimators can then be formed by evaluating the previous expression at the MLE, that is $\widehat{\mathcal{B}} = \widehat{\mathcal{B}}(\widehat{\theta})$ and $\tilde{\theta} = \widehat{\theta} - \widehat{\mathcal{B}}/T$, and iterated bias corrected estimator can be obtained as the solution to $\tilde{\theta}^{\infty} = \widehat{\theta} - \widehat{\mathcal{B}}(\tilde{\theta}^{\infty})/T$. A score-corrected estimator can be also obtained by solving the modified first order condition:

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{E}_T\left[\widehat{u}_{it}(\tilde{\theta}^{sc})\right] - \frac{1}{T}\widehat{b}(\tilde{\theta}^{sc}) = 0, \tag{B.7}$$

where $\widehat{u}_{it}(\theta) = \widehat{H}_{it}(\theta)\left[Y_{it} - \widehat{F}_{it}(\theta)\right]X_{it}$.

B.2 Bias Corrections for Average Marginal effects

Let $m(x_{it}, \theta, \alpha_i)$ denote generically one of the measures for the individual marginal effects described in Section 2, i.e., $m(x_{it}, \theta, \alpha)$ or $\tilde{m}(x_{it}, \theta, \alpha_i)$. The object of interest is then $\mu = E_n\left[E_T\left[m(x_{it}, \theta_0, \alpha)\right]\right]$, and a fixed effects estimator can be constructed as

$$\widehat{\mu}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\widehat{E}_T\left[m\left(x_{it}, \theta, \widehat{\alpha}_i(\theta)\right)\right]. \tag{B.8}$$

For the bias corrections, let $\tilde{\theta}$ be a bias-corrected estimator of θ_0 and $\tilde{\alpha}_i = \widehat{\alpha}_i(\tilde{\theta})$, $i = 1, ..., n$, the corresponding estimators of the individual effects. Then, large-T consistent estimators for the individual components of the bias in equation (4.5) can be formed by replacing population moments for sample analogs and true parameter values for bias corrected estimates, that is

$$\widehat{\Delta}_i = \widehat{E}_T\left[m_\alpha\left(x_{it}, \tilde{\theta}, \tilde{\alpha}_i\right)\right]\widehat{\beta}_i\left(\tilde{\theta}\right) + \widehat{\widetilde{E}}_{T,J}\left[m_\alpha\left(x_{it}, \tilde{\theta}, \tilde{\alpha}_i\right)\widehat{\psi}_{is}\left(\tilde{\theta}\right)\right] + \frac{1}{2}\widehat{E}_T\left[m_{\alpha\alpha}\left(x_{it}, \tilde{\theta}, \tilde{\alpha}_i\right)\right]\widehat{\sigma}_i^2\left(\tilde{\theta}\right), \tag{B.9}$$

where subscripts on m denote partial derivatives, and J is a bandwidth parameter chosen such that $J/T^{1/2} \to 0$ as $T \to \infty$. The bias corrected estimator for μ can be formed as

$$\tilde{\mu} = \widehat{\mu}(\tilde{\theta}) - \frac{1}{nT}\sum_{i=1}^{n}\widehat{\Delta}_i. \tag{B.10}$$

C Proofs

C.1 Lemmas

Lemma 1 ***(McFadden and Reid, 1975)*** *Let $Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$, and $a, b \in \mathbb{R}$ with $b > 0$. Then,*

$$\Phi\left(\frac{\mu_Z + a}{\sqrt{b^2 + \sigma_Z^2}}\right) = \int \Phi\left(\frac{z+a}{b}\right) \frac{1}{\sigma_Z}\ \phi\left(\frac{z-\mu_Z}{\sigma_Z}\right) dz, \tag{C.1}$$

and

$$\frac{1}{\sqrt{b^2 + \sigma_Z^2}}\ \phi\left(\frac{\mu_Z + a}{\sqrt{b^2 + \sigma_Z^2}}\right) = \int \frac{1}{b}\ \phi\left(\frac{z+a}{b}\right) \frac{1}{\sigma_Z}\ \phi\left(\frac{z-\mu_Z}{\sigma_Z}\right) dz, \tag{C.2}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote cdf and pdf of the standard normal distribution, respectively.

Proof. First, take X independent of Z, with $X \sim \mathcal{N}(-a, b^2)$. Then,

$$\Pr\{X - Z \leq 0\} = \Phi\left(\frac{\mu_Z + a}{\sqrt{b^2 + \sigma_Z^2}}\right) \tag{C.3}$$

since $X - Z \sim \mathcal{N}(-a - \mu_Z, b^2 + \sigma_Z^2)$. Alternatively, using the law of iterated expectations and $X|Z \sim X$ by independence,

$$\Pr\{X - Z \leq 0\} = E_Z\left[\Pr\{X \leq Z|Z\}\right] = \int \Phi\left(\frac{z+a}{b}\right) \frac{1}{\sigma_Z} \phi\left(\frac{z-\mu_Z}{\sigma_Z}\right) dz. \tag{C.4}$$

The second statement follows immediately by deriving both sides of expression (C.1) with respect to a. ■

C.2 Proof of Proposition 1

Proof. First, note that for the probit $g_{it} = -(X_{it}'\theta_0 + \alpha_i) f_{it}$. Then, substituting this expression for g_{it} in the bias formulas of the static model and dropping the terms that involve $\tilde{E}_T$, see Appendix A, yields

$$\beta_i = \left\{\sigma_i^4 E_T\left[H_{it} f_{it} X_{it}'\right] \theta_0 + \sigma_i^2 \alpha_i\right\}/2, \tag{C.5}$$

$$\mathcal{J} = -E_n\left\{E_T\left[H_{it} f_{it} X_{it} X_{it}'\right] - \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}\right] E_T\left[H_{it} f_{it} X_{it}'\right]\right\} = E_n\left[\mathcal{J}_i\right], \tag{C.6}$$

$$\begin{aligned} b &= -E_n\left\{\sigma_i^4 E_T\left[H_{it} f_{it} X_{it}\right] E_T\left[H_{it} f_{it} X_{it}'\right] \theta_0 + \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}\right] \alpha_i\right\}/2 \\ &+ E_n\left\{\sigma_i^2 E_T\left[H_{it} f_{it} X_{it} X_{it}'\right] \theta_0 + \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}\right] \alpha_i\right\}/2 \\ &= E_n\left\{\sigma_i^2\left(E_T\left[H_{it} f_{it} X_{it} X_{it}'\right] - \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}\right] E_T\left[H_{it} f_{it} X_{it}'\right]\right)\right\} \theta_0/2 = -E_n\left[\sigma_i^2 \mathcal{J}_i\right] \theta_0/2. \end{aligned} \tag{C.7}$$

Finally, we have for the bias

$$\mathcal{B} = -\mathcal{J}^{-1} b = \frac{1}{2} E_n\left[\mathcal{J}_i\right]^{-1} E_n\left[\sigma_i^2 \mathcal{J}_i\right] \theta_0. \tag{C.8}$$

The second and third results are immediate, and are described in the text. ■

C.3 Proof of Proposition 2

Proof. We want to find the probability limit of $\widehat{\mu} = \sum_{i=1}^{n} \widehat{\theta}\ \widehat{E}_T \left[\phi\left(X_{it}'\widehat{\theta} + \widehat{\alpha}_i(\widehat{\theta})\right)\right]/n$, as $n, T \to \infty$, and compare it to the population parameter of interest $\mu = E_n \left[\theta_0 \phi\left(X_{it}'\theta_0 + \alpha_i\right)\right]$.

First, note that by the Law of Large Number and Continuous Mapping Theorem, as $n, T \to \infty$

$$\widehat{\mu} \xrightarrow{p} E_n \left\{\theta_T E_T \left[\phi\left(X_{it}'\theta_T + \widehat{\alpha}_i(\theta_T)\right)\right]\right\}, \tag{C.9}$$

where $\bar{\theta}$ lies between θ_T and θ_0. Next, we have the following expansion for the limit index, $\widehat{\xi}_{it}(\theta_T) := X_{it}'\theta_T + \widehat{\alpha}_i(\theta_T)$, around θ_0

$$\widehat{\xi}_{it}(\theta_T) \quad = \quad X_{it}'\theta_0 + \widehat{\alpha}_i(\theta_0) + \left[X_{it}' + \frac{\partial \widehat{\alpha}_i(\bar{\theta})}{\partial \theta'}\right](\theta_T - \theta_0). \tag{C.10}$$

Using independence across t, standard higher-order asymptotics for $\widehat{\alpha}_i(\theta_0)$ give (e.g., Ferguson, 1992, or Rilstone *et al.*, 1996), as $T \to \infty$

$$\widehat{\alpha}_i(\theta_0) = \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + R_{1i}, \quad \psi_i \xrightarrow{d} \mathcal{N}(0, \sigma_i^2 := -E_T\left[v_{it\alpha}\right]^{-1}), \tag{C.11}$$

where $R_{1i} = O_p(T^{-3/2})$ and $E_T\left[R_{1i}\right] = O(T^{-2})$ uniformly in i by the conditions of the proposition (e.g., HN and Fernández-Val, 2004). From the first order conditions for $\widehat{\alpha}_i(\theta)$, we have, as $n, T \to \infty$

$$\frac{\partial \widehat{\alpha}_i(\bar{\theta})}{\partial \theta'} = -\frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]} + R_{2i} = \sigma_i^2 E_T\left[v_{it\theta}\right] + R_{2i}, \tag{C.12}$$

where $R_{2i} = O_p(T^{-1/2})$ and $E_T\left[R_{2i}\right] = O(T^{-1})$ uniformly in i, again by the conditions of the proposition. Plugging (C.11) and (C.12) into the expansion for the index in (C.10) yields, for $\xi_{it} = X_{it}'\theta_0 + \alpha_i$,

$$\widehat{\xi}_{it}(\theta_T) \quad = \quad \xi_{it} + \psi_i/\sqrt{T} + \beta_{\xi_i}/T + R_{3i}, \tag{C.13}$$

where $\beta_{\xi_i} = \beta_i + T\left(X_{it}' + \sigma_i^2 E_T\left[v_{it\theta}\right]\right)(\theta_T - \theta_0)$, $R_{3i} = O_p(T^{-3/2})$ and $E_T\left[R_{3i}\right] = O(T^{-2})$, uniformly in i by the properties of R_{1i} and R_{2i}.

Then, using the expressions for the bias for the static probit model, see proof of Proposition 1, and $E_T\left[v_{it\theta}\right] = -E_T\left[H_{it} f_{it} X_{it}'\right]$, we have

$$\begin{aligned}
\beta_{\xi_i} \quad &= \quad \left(\sigma_i^4 E_T\left[H_{it} f_{it} X_{it}'\right]\theta_0 + \sigma_i^2 \alpha_i\right)/2 + \left(X_{it}' - \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}'\right]\right)\mathcal{B} + O(T^{-2}) \\
&= \quad \sigma_i^2 \xi_{it}/2 - \left(X_{it}' - \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}'\right]\right)\sigma_i^2\theta_0/2 + \left(X_{it}' - \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}'\right]\right)E_n\left[\mathcal{J}_i\right]^{-1} E_n\left[\sigma_i^2 \mathcal{J}_i\right]\theta_0 \\
&+ \quad O(T^{-2}) = \sigma_i^2 \xi_{it}/2 - \mathcal{D}_i + O(T^{-2}),
\end{aligned} \tag{C.14}$$

where $\mathcal{D}_i = \left(X_{it}' - \sigma_i^2 E_T\left[H_{it} f_{it} X_{it}'\right]\right)\left(\sigma_i^2 \mathcal{I}_p - E_n\left[\mathcal{J}_i\right]^{-1} E_n\left[\sigma_i^2 \mathcal{J}_i\right]\right)\theta_0/2$, $\mathcal{I}_p$ denotes the $p \times p$ identity matrix, and the remainder term is uniformly bounded in i. Substituting the expression for β_{ξ_i} in (C.13) gives

$$\widehat{\xi}_{it}(\theta_T) = \left[1 + \sigma_i^2/2T\right]\xi_{it} + \psi_i/\sqrt{T} - \mathcal{D}_i/T + R_i, \quad \psi_i/\sqrt{T} \overset{a}{\sim} \mathcal{N}(0, \sigma_i^2/T) \tag{C.15}$$

where $R_i = O_p(T^{-3/2})$ and $E_T\left[R_i\right] = O(T^{-2})$ uniformly in i.

Finally, using Lemma 1 and expanding around θ_0, it follows that, as $n, T \to \infty$,

$$
\begin{aligned}
\widehat{\mu} &\xrightarrow{p} E_n \left\{ \theta_T E_T \left[\phi \left(\widehat{\xi}_i(\theta_T) \right) \right] \right\} \\
&= E_n \left\{ \theta_T \int \phi \left(\left[1 + \sigma_i^2/2T \right] \xi_{it} + v - \mathcal{D}_i/T \right) \frac{\sqrt{T}}{\sigma_i} \, \phi \left(\frac{\sqrt{T} v}{\sigma_i} \right) dv + O(T^{-2}) \right\} \\
&= E_n \left\{ \left(1 + \sigma_i^2/T \right)^{-1/2} \theta_T \phi \left(\frac{\left[1 + \sigma_i^2/2T \right] \xi_{it} - \mathcal{D}_i/T}{\sqrt{1 + \sigma_i^2/T}} \right) + O(T^{-2}) \right\} \\
&= \mu + \frac{1}{2T} E_n \left\{ \phi(\xi_{it}) \left(\xi_{it} \theta_0 \left(X_{it} - \sigma_i^2 E_T \left[H_{it} f_{it} X_{it} \right] \right)' - \mathcal{I}_p \right) \left(\sigma_i^2 \mathcal{I}_p - E_n \left[\mathcal{J}_i \right]^{-1} E_n \left[\sigma_i^2 \mathcal{J}_i \right] \right) \theta_0 \right\} \\
&+ O(T^{-2}) = \mu + \frac{1}{T} \mathcal{B}_\mu + O\left(T^{-2} \right), \qquad \text{(C.16)}
\end{aligned}
$$

where the remainder terms are uniformly bounded in i because the first three derivatives of $\phi(\xi_{it})$ are bounded by the conditions of the proposition, and

$$
\frac{\sigma_i^2}{\sqrt{1 + \sigma_i^2/T}} = \left(1 + \frac{\sigma_i^2}{2T} \right) \left(1 - \frac{\sigma_i^2}{2T} + O(T^{-2}) \right) = 1 + O(T^{-2}). \qquad \text{(C.17)}
$$

■

References

[1] Alvarez, J., and M. Arellano (2003), "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica* 71(4), 1121-1159.

[2] Amemiya, T. (1981), "Qualitative Response Models: A Survey," *Journal of Economic Literature* 19(4), 1483-1536.

[3] Andersen, E. B. (1973), *Conditional Inference and Models for Measuring.* Mentalhygiejnisk Forlag. Copenhagen.

[4] Angrist, J. D. (2001), "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics* 19(1), 2-16.

[5] Angrist, J. D., and W. H. Evans (1998), "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review* 88(3), 450-477.

[6] Arellano, M. (2003), "Discrete Choices with Panel Data," *Investigaciones Económicas* XXVII (3), 423-458.

[7] Arellano, M., and R. Carrasco (2003), "Discrete Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics* 115 (1), 125-157.

[8] Arellano, M., and B. Honoré (2001), "Panel Data Models: Some Recent Developments," in J. J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Vol. 5*, Amsterdam: North-Holland.

[9] Card, D., and D. Sullivan (1988), "Measuring the Effect of Subsidied Training Programs on Movements In and Out of Employment," *Econometrica* 56, 497-530.

[10] Carneiro, P., Heckman, J., and K. Hansen (2003), "Estimating Distribution of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effect of Uncertainty on Schooling Choice," *International Economic Review* 44(2), pp 361-422

[11] Carro, J. M. (2006), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," " *Journal of Econometrics*, doi: 1016/j-jeconom.2006.07.023.

[12] Carrasco, R. (2001), "Binary Choice With Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation," *Journal of Business and Economic Statistics* 19(4), 385-394.

[13] Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* XLVII, 225-238.

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

[15] Chamberlain, G. (1985), "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in J. J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*. Cambridge University Press.

[16] Chamberlain, G., and E. E. Leamer (1976), "Matrix Weighted Averages and Posterior Bounds," *Journal of the Royal Statistical Society. Series B* 38(1), 73-84.

[17] Chay, K. Y., and D. R. Hyslop (2000), "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches," unpublished manuscript, University of California at Berkeley.

[18] Chesher, A., and I. Jewitt (1987), "The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator," *Econometrica* 55, 1217-1222.

[19] Cox, D. R. (1958), "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society. Series B* 20(2), 215-242.

[20] Fernández-Val, I. (2004), "Bias Correction in Panel Data Models with Individual Specific Parameters," working paper, MIT Department of Economics.

[21] Fernández-Val, I. (2005), "Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects," working paper, MIT Department of Economics.

[22] Ferguson, H. (1992), "Asymptotic Properties of a Conditional Maximum-Lkelihood Estimator," *The Canadian Journal of Statistics* 20(1), 63-75.

[23] Greene, W.H. (2002), "The Behavior of the Fixed Effects Estimator in Nonlinear Models,"unpublished manuscript, New York University.

[24] Hahn, J. (2001), "Comment: Binary Regressors in Nonlinear Panel-Data Models With Fixed Effects," *Journal of Business and Economic Statistics* 19(1), 16-17.

[25] Hahn, J., and G. Kuersteiner (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica* 70, 1639-1657.

[26] Hahn, J., and G. Kuersteiner (2003), "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," unpublished manuscript. UCLA.

[27] Hahn, J., and W. Newey (2004), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica* 72, 1295-1319.

[28] Heckman, J. J. (1981a), "Statistical Models for Discrete Data," in C. F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Panel Data with Econometric Applications*, 114-178.

[29] Heckman, J. J. (1981b), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process, " in C. F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Panel Data with Econometric Applications*, 179-195.

[30] Heckman, J. J., and T. E. MaCurdy (1980), "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies* 47, 47-74.

[31] Heckman, J. J., and T. E. MaCurdy (1982), "Corrigendum on: A Life Cycle Model of Female Labor Supply," *Review of Economic Studies* 49, 659-660.

[32] Honoré, B. E., and E. Kyriazidou (2000), "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 839-874.

[33] Honoré, B. E., and A. Lewbel (2002), "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica* 70, 2053-2063.

[34] Hyslop, D. R. (1999), "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica* 67(6), 1255-1294.

[35] Katz, E. (2001), "Bias in Conditional and Unconditional Fixed Effects Logit Estimation," *Political Analysis* 9(4), 379-384.

[36] Lancaster, T. (2000), "The Incidental Parameters Problem Since 1948," *Journal of Econometrics* 95, 391-413.

[37] Lancaster, T. (2002), "Orthogonal Parameters and Panel Data," *Review of Economic Studies* 69, 647-666.

[38] Lewbel, A. (2005), "Simple Endogenous Binary Choice and Selection Panel Model Estimators," unpublished manuscript, Boston College.

[39] Manski, C. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica* 55, 357-362.

[40] Moon, C.-G., and J. G. Stotsky (1993), "The Effect of Rent Control on Housing Quality Change: A Longitudinal Analysis," *Journal of Political Economy* 101, 1114-1148.

[41] Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica* 46, 69-85.

[42] Nerlove, M. (1967), "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," *Economic Studies Quarterly* 18, 42-74.

[43] Nerlove, M. (1971), "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," *Econometrica* 39, 359-387.

[44] Newey, W. K., and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Vol. 4*. Elsevier Science. Amsterdam: North-Holland.

[45] Neyman, J., and E.L. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16, 1-32.

[46] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica* 49, 1417-1426.

[47] Phillips, P. C. B., and H. R. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* 67, 1057-1111.

[48] Porter, J. (2002), "Efficiency of Covariance Matrix Estimators for Maximum Likelihood Estimation," *Review of Business and Economic Statistics* 20(3), 431-440.

[49] Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, 2nd Edition. New York: John Wiley.

[50] Rilstone, P., V. K. Srivastava, and A. Ullah (1996), "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators, " *Journal of Econometrics* 75, 369-395.

[51] Roberts, M. J., and J. R. Tybout (1997), "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs, " *The American Economic Review* 87(4), 545-564.

[52] Rudd, P. A. (2000), *An Introduction to Classical Econometric Theory*, Oxford University Press, New York.

[53] Stoker, T. M. (1986), "Consistent Estimation of Scaled Coefficients," *Econometrica* 54, 1461-1481.

[54] Wooldridge, J. (2001), "A Framework for Estimating Dynamic, Unobserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables," *Economics Letters* 68, 245-250.

[55] Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

[56] Wooldridge, J. M. (2005), "Unobserved Heterogeneity and Estimation of Average Partial Effects," in D.W.K. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometrics Models: A Festschrift in Honor of Thomas Rothenberg*, Cambridge, Cambridge University Press.

[57] Woutersen, T. M. (2002), "Robustness Against Incidental Parameters," unpublished manuscript, University of Western Ontario.

[58] Yatchew, A., and Z. Griliches (1985), "Specification Error in Probit Models," *Review of Economics and Statistics* 67, 134-139.

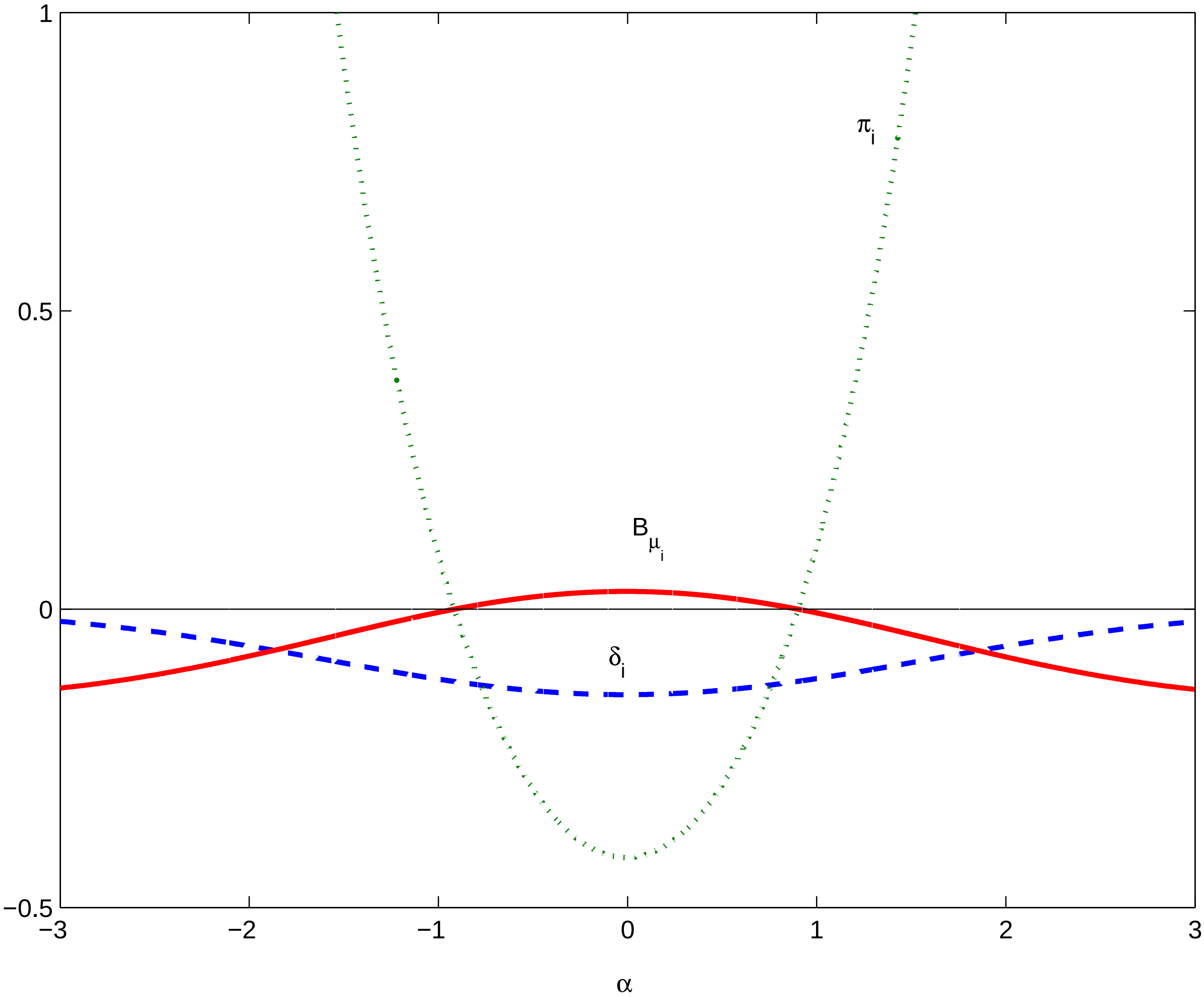

Figure 1: Components of the bias of the fixed effects estimator of the marginal effect: $\mathcal{B}_\mu = E_\alpha[\mathcal{B}_{\mu_i}] = E_\alpha[\delta_i \pi_i]$. Individual effects and regressor generated from independent standard normal distributions.

Table 1: True Bias (in parenthesis) and First Order Approximation, T = 8

	Regressor				
Individual Effects	Nerlove	Normal	$\chi^2(1)$	$\chi^2(2)$	Bi(10,.9)
	A - Index Coefficients (% true parameter value)				
Normal	14.59	15.62	16.56	15.93	15.62
	(18.72)	(21.06)	(15.57)	(19.07)	(21.98)
$\chi^2(1)$	12.57	14.77	13.82	14.00	14.85
	(16.00)	(19.86)	(15.32)	(19.33)	(19.83)
$\chi^2(2)$	13.17	14.95	13.04	14.55	15.16
	(16.50)	(20.91)	(15.93)	(19.64)	(20.19)
Bi(10,.9)	15.46	15.53	16.36	16.47	15.44
	(19.52)	(21.32)	(15.16)	(19.62)	(21.77)
C = 0	12.87	13.22	11.91	12.39	13.17
	(17.12)	(17.72)	(11.76)	(15.55)	(17.60)
	B - Marginal Effects (% true parameter value)				
Normal	-0.27	-0.07	1.89	0.95	-0.03
	(-0.45)	(-0.40)	(-1.92)	(-0.47)	(0.19)
$\chi^2(1)$	-0.21	-0.06	-0.18	-0.24	0.34
	(-0.10)	(0.08)	(-1.50)	(0.23)	(0.38)
$\chi^2(2)$	-0.28	-0.05	-0.54	-0.10	0.31
	(-0.04)	(0.24)	(-1.49)	(-0.07)	(0.15)
Bi(10,.9)	-0.27	-0.08	1.56	1.59	-0.26
	(-0.56)	(-0.35)	(-1.85)	(0.08)	(-0.09)
C = 0	0.00	0.00	0.00	0.00	0.00
	(-0.44)	(0.53)	(-1.84)	(0.12)	(0.50)

Notes: Bias formulae for first order approximation evaluated numerically using 10,000 replications. True bias obtained by Monte Carlo with 1,000 repetitions.

Table 2: Estimators of θ ($\theta_0 = 1$), ε ~ N(0,1)

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 4				
PROBIT	1.41	1.40	0.393	0.25	0.36	0.82	0.410
JK-PROBIT	0.75	0.75	0.277	0.11	0.19	1.08	0.265
BC1-PROBIT	1.11	1.10	0.304	0.04	0.11	1.03	0.215
BC2-PROBIT	1.20	1.19	0.333	0.09	0.16	0.95	0.253
BC3-PROBIT	1.06	1.06	0.275	0.02	0.06	1.13	0.195
			T = 8				
PROBIT	1.18	1.18	0.151	0.28	0.37	0.90	0.180
JK-PROBIT	0.95	0.96	0.118	0.05	0.11	1.09	0.085
BC1-PROBIT	1.05	1.05	0.134	0.05	0.11	0.98	0.099
BC2-PROBIT	1.05	1.05	0.132	0.05	0.10	1.00	0.097
BC3-PROBIT	1.02	1.02	0.124	0.03	0.07	1.05	0.085
			T = 12				
PROBIT	1.13	1.13	0.096	0.30	0.41	0.94	0.129
JK-PROBIT	0.98	0.98	0.080	0.05	0.10	1.06	0.055
BC1-PROBIT	1.04	1.04	0.087	0.07	0.13	0.99	0.062
BC2-PROBIT	1.03	1.03	0.085	0.06	0.11	1.01	0.058
BC3-PROBIT	1.01	1.01	0.082	0.04	0.09	1.05	0.056

Notes: 1,000 replications. JK denotes Hahn and Newey (2004) Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey (2004) bias-corrected estimator based on general estimating equations; BC3 denotes the bias-corrected estimator proposed in the paper.

Table 3: Estimators of μ (true value = 1), ε ~ N(0,1)

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 4				
PROBIT	0.99	0.99	0.242	0.10	0.16	0.82	0.163
JK-PROBIT	1.02	1.02	0.285	0.12	0.19	0.75	0.182
BC1-PROBIT	1.00	1.00	0.261	0.12	0.18	0.79	0.176
BC2-PROBIT	1.04	1.04	0.255	0.12	0.19	0.80	0.176
BC3-PROBIT	0.94	0.94	0.226	0.08	0.13	0.91	0.158
LPM	0.98	0.98	0.233	0.09	0.15	0.84	0.156
LPM-FS	1.00	1.00	0.242	0.10	0.16	0.87	0.163
			T = 8				
PROBIT	0.99	0.99	0.104	0.08	0.14	0.82	0.070
JK-PROBIT	1.00	1.00	0.107	0.07	0.14	0.84	0.071
BC1-PROBIT	1.01	1.01	0.110	0.09	0.15	0.80	0.073
BC2-PROBIT	1.00	1.00	0.105	0.07	0.13	0.83	0.070
BC3-PROBIT	0.97	0.97	0.103	0.08	0.13	0.86	0.071
LPM	0.98	0.98	0.104	0.07	0.14	0.84	0.071
LPM-FS	1.00	1.00	0.109	0.07	0.13	0.87	0.075
			T = 12				
PROBIT	0.99	0.99	0.062	0.05	0.11	0.75	0.043
JK-PROBIT	1.00	1.00	0.064	0.05	0.11	0.76	0.042
BC1-PROBIT	1.00	1.00	0.065	0.06	0.11	0.74	0.042
BC2-PROBIT	0.99	0.99	0.062	0.05	0.10	0.76	0.042
BC3-PROBIT	0.98	0.98	0.062	0.05	0.11	0.77	0.043
LPM	0.99	0.99	0.065	0.06	0.11	0.76	0.041
LPM-FS	1.01	1.01	0.067	0.05	0.11	0.80	0.045

Notes: 1,000 replications. JK denotes Hahn and Newey (2004) Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey (2004) bias-corrected estimator based on general estimating equations; BC3 denotes the bias-corrected estimator proposed in the paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model.

Table 4: Estimators of θ_Y ($\theta_{Y,0} = .5$), $\varepsilon \sim L(0,\pi^2/3)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
LOGIT	-0.26	-0.26	0.185	0.99	1.00	0.92	0.760
BC1-LOGIT	0.38	0.38	0.161	0.14	0.21	0.95	0.143
HK-LOGIT		0.45					0.131
MML-LOGIT		0.39		0.11			0.127
BC3-LOGIT	0.44	0.43	0.148	0.07	0.13	0.98	0.112
			T = 12				
LOGIT	0.07	0.06	0.123	0.95	0.97	0.99	0.435
BC1-LOGIT	0.45	0.45	0.110	0.07	0.13	1.03	0.084
BC3-LOGIT	0.47	0.47	0.108	0.05	0.10	1.03	0.073
			T = 16				
LOGIT	0.19	0.18	0.101	0.88	0.93	0.98	0.315
BC1-LOGIT	0.46	0.46	0.092	0.07	0.12	1.03	0.070
HK-LOGIT		0.45					0.074
MML-LOGIT		0.48					0.067
BC3-LOGIT	0.48	0.47	0.093	0.05	0.11	1.01	0.067

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 5: Estimators of θ_X ($\theta_{X,0} = 1$), $\varepsilon \sim L(0,\pi^2/3)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
LOGIT	1.26	1.26	0.081	0.95	0.98	0.92	0.255
BC1-LOGIT	1.09	1.09	0.073	0.34	0.46	0.81	0.090
HK-LOGIT		1.01					0.050
MML-LOGIT		1.01		0.06			0.039
BC3-LOGIT	0.99	0.99	0.053	0.07	0.12	0.97	0.037
			T = 12				
LOGIT	1.15	1.15	0.054	0.83	0.90	0.93	0.146
BC1-LOGIT	1.04	1.04	0.049	0.15	0.22	0.88	0.042
BC3-LOGIT	1.00	1.00	0.044	0.07	0.11	0.93	0.030
			T = 16				
LOGIT	1.10	1.10	0.041	0.71	0.81	0.99	0.100
BC1-LOGIT	1.02	1.02	0.038	0.08	0.15	0.96	0.027
HK-LOGIT		1.01					0.029
MML-LOGIT		1.01					0.023
BC3-LOGIT	1.00	1.00	0.036	0.05	0.10	0.98	0.024

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 6: Estimators of μ_Y (true value = 1), $\varepsilon \sim L(0,\pi^2/3)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
LOGIT	-0.39	-0.40	0.277	1.00	1.00	0.92	1.396
BC1-LOGIT	0.71	0.70	0.304	0.25	0.34	0.85	0.319
BC3-LOGIT	0.86	0.86	0.300	0.13	0.19	0.86	0.226
LPM	-0.45	-0.46	0.276	1.00	1.00	0.95	1.456
BC-LPM	0.76	0.76	0.306	0.18	0.27	0.87	0.281
LPM-FS	-0.54	-0.55	0.303	1.00	1.00	0.96	1.554
BC-LPM-FS	0.84	0.84	0.336	0.12	0.19	0.87	0.262
			T = 12				
LOGIT	0.11	0.11	0.210	0.99	0.99	0.99	0.890
BC1-LOGIT	0.87	0.87	0.217	0.11	0.17	0.96	0.175
BC3-LOGIT	0.95	0.94	0.220	0.07	0.13	0.95	0.153
LPM	0.06	0.06	0.218	0.99	1.00	1.00	0.940
BC-LPM	0.91	0.91	0.232	0.08	0.13	0.95	0.166
LPM-FS	0.05	0.05	0.224	0.99	0.99	1.00	0.947
BC-LPM-FS	0.94	0.94	0.240	0.08	0.12	0.94	0.164
			T = 16				
LOGIT	0.34	0.33	0.183	0.95	0.97	0.97	0.669
BC1-LOGIT	0.91	0.90	0.185	0.09	0.17	0.97	0.146
BC3-LOGIT	0.95	0.94	0.188	0.07	0.14	0.95	0.134
LPM	0.30	0.30	0.191	0.96	0.98	0.99	0.701
BC-LPM	0.95	0.94	0.200	0.08	0.13	0.94	0.142
LPM-FS	0.30	0.30	0.193	0.95	0.98	0.99	0.700
BC-LPM-FS	0.96	0.95	0.203	0.07	0.13	0.94	0.138

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model; BC-LPM denotes Nickell (1981) bias-corrected adjusted linear probability model; BC-LPM-FS denotes Nickell (1981) bias-corrected linear probability model. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 7: Estimators of μ_X (true value = 1), $\varepsilon \sim L(0,\pi^2/3)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
LOGIT	0.98	0.98	0.035	0.07	0.12	0.90	0.028
BC1-LOGIT	1.01	1.01	0.040	0.12	0.19	0.75	0.028
BC3-LOGIT	0.98	0.98	0.034	0.11	0.18	0.81	0.028
LPM	0.95	0.95	0.033	0.35	0.47	0.86	0.049
BC-LPM	0.97	0.97	0.033	0.20	0.29	0.86	0.035
LPM-FS	0.97	0.97	0.033	0.12	0.20	0.95	0.030
BC-LPM-FS	0.99	0.99	0.033	0.06	0.11	0.95	0.025
			T = 12				
LOGIT	1.00	1.00	0.025	0.03	0.07	0.92	0.017
BC1-LOGIT	1.01	1.01	0.026	0.06	0.11	0.84	0.019
BC3-LOGIT	1.00	1.00	0.025	0.04	0.09	0.86	0.017
LPM	0.98	0.98	0.025	0.12	0.19	0.89	0.021
BC-LPM	0.99	0.99	0.025	0.08	0.14	0.89	0.018
LPM-FS	0.99	0.99	0.026	0.09	0.14	0.91	0.019
BC-LPM-FS	0.99	0.99	0.026	0.06	0.12	0.91	0.018
			T = 16				
LOGIT	1.00	1.00	0.020	0.02	0.07	0.94	0.013
BC1-LOGIT	1.00	1.00	0.020	0.04	0.09	0.89	0.013
BC3-LOGIT	1.00	1.00	0.020	0.03	0.08	0.90	0.013
LPM	0.99	0.99	0.021	0.08	0.14	0.92	0.016
BC-LPM	0.99	0.99	0.021	0.07	0.12	0.92	0.015
LPM-FS	0.99	0.99	0.021	0.07	0.13	0.93	0.015
BC-LPM-FS	0.99	1.00	0.021	0.06	0.11	0.93	0.015

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model; BC-LPM denotes Nickell (1981) bias-corrected adjusted linear probability model; BC-LPM-FS denotes Nickell (1981) bias-corrected linear probability model. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 8: Descriptive Statistics, Married Women (n = 1461, T = 9)

	Full Sample		Always Participate		Never Participate		Movers		
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Within (%)
Participation	0.72	0.45	1	0	0	0	0.57	0.49	72.51
Age (in 1985)	37.30	9.22	37.98	9.04	42.98	10.09	35.57	8.71	9.09
Black	0.21	0.40	0.24	0.43	0.25	0.43	0.16	0.37	0
Years of schooling	12.26	3.79	12.49	3.88	12.09	3.20	12.05	3.77	0
Kids 0-2	0.23	0.47	0.18	0.41	0.21	0.47	0.28	0.51	62.96
Kids 3-5	0.29	0.51	0.23	0.46	0.23	0.48	0.36	0.56	65.52
Kids 6-17	1.05	1.10	1.00	1.06	0.99	1.19	1.11	1.11	33.67
Husband income (1995 $1000)	42.29	40.01	38.33	25.15	53.27	74.62	44.32	42.69	20.99
No. Observations	13149		6084		1089		5976		

Source: PSID 1980-1988.

Table 9: Female Labor Force Participation (n = 1461, T = 9), Static Model

	PROBIT				LOGIT					LPM			
Estimator	FE	JK	BC3	BC3p	FE	JK	BC3	C	BC3p	FE	FE-FS	BC	BC-FS
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
A - Index Coefficients													
Kids 0-2	-0.71	-0.61	-0.63	-0.66	-0.68	-0.59	-0.60	-0.60	-0.63				
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)				
Kids 3-5	-0.42	-0.37	-0.37	-0.39	-0.40	-0.35	-0.35	-0.35	-0.37				
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)				
Kids 6-17	-0.13	-0.10	-0.11	-0.13	-0.13	-0.11	-0.11	-0.11	-0.13				
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
Log(Husband income)	-0.25	-0.22	-0.22	-0.22	-0.24	-0.21	-0.21	-0.21	-0.21				
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)				
B - Marginal Effects (%)													
Kids 0-2	-9.22	-9.38	-9.07	-9.59	-9.35	-9.35	-9.20		-9.73	-9.46	-11.19	-9.92	-11.72
	(0.70)	(0.71)	(0.70)	(0.77)	(0.72)	(0.72)	(0.72)		(0.76)	(0.75)	(0.88)	(0.80)	(0.94)
Kids 3-5	-5.45	-5.60	-5.36	-5.63	-5.53	-5.59	-5.45		-5.72	-5.54	-6.09	-5.80	-6.38
	(0.66)	(0.66)	(0.66)	(0.71)	(0.67)	(0.67)	(0.67)		(0.71)	(0.70)	(0.81)	(0.75)	(0.88)
Kids 6-17	-1.68	-1.59	-1.66	-1.85	-1.78	-1.72	-1.76		-1.94	-1.78	-1.25	-1.97	-1.45
	(0.53)	(0.53)	(0.53)	(0.59)	(0.54)	(0.54)	(0.54)		(0.59)	(0.58)	(0.56)	(0.63)	(0.61)
Log(Husband income)	-3.25	-3.31	-3.20	-3.15	-3.26	-3.29	-3.22		-3.17	-3.17	-3.61	-3.09	-3.59
	(0.70)	(0.69)	(0.70)	(0.74)	(0.71)	(0.71)	(0.71)		(0.75)	(0.72)	(0.71)	(0.74)	(0.74)

Notes: All the specifications include time dummies and a quadratic function of age. FE denotes uncorrected fixed effects estimator; JK denotes Hahn and Newey (2004) Jackknife bias-corrected estimator; BC3 denotes the bias-corrected estimator proposed in the paper; BC3p denotes the bias-corrected estimator proposed in the paper when the regressors are treated as predetermined; C denotes conditional logit estimator; LPM-FE denotes adjusted linear probability model (see text); LPM-FE-FS denotes linear probability model; LPM-BC denotes a bias corrected adjusted linear probability model for predetermined regressors; LPM-BC-FS denotes a bias corrected linear probability model. Logit estimates and standard errors of index coefficients are normalized to have the same scale as probit.
Source: PSID 1980-1988.

Table 10: Female Labor Force Participation (n = 1461, T = 10), Dynamic Model

	PROBIT			LOGIT		LPM			
Estimator	FE	BC3	MML	FE	BC3	FE	BC	FE-FS	BC-FS
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
				A - Index Coefficients					
$\text{Participation}_{t-1}$	0.76	1.03	1.08	0.69	0.95				
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
Kids 0-2	-0.55	-0.44	-0.40	-0.53	-0.42				
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)				
Kids 3-5	-0.29	-0.20	-0.18	-0.27	-0.19				
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)				
Kids 6-17	-0.07	-0.05	-0.04	-0.07	-0.05				
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)				
Log(Husband income)	-0.25	-0.21	-0.21	-0.24	-0.20				
	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)				
				B - Marginal Effects (%)					
$\text{Participation}_{t-1}$	10.69	17.08		10.47	17.18	11.42	16.17	25.58	35.66
	(0.64)	(0.67)		(0.64)	(0.66)	(0.63)	(0.64)	(1.30)	(1.33)
Kids 0-2	-6.76	-5.94		-6.81	-5.99	-6.87	-6.24	-8.02	-7.29
	(0.74)	(0.73)		(0.74)	(0.72)	(0.75)	(0.73)	(0.86)	(0.85)
Kids 3-5	-3.55	-2.77		-3.53	-2.72	-3.44	-2.78	-3.57	-2.81
	(0.68)	(0.67)		(0.68)	(0.67)	(0.69)	(0.67)	(0.79)	(0.77)
Kids 6-17	-0.91	-0.67		-0.95	-0.71	-0.93	-0.74	-0.49	-0.38
	(0.54)	(0.54)		(0.54)	(0.53)	(0.55)	(0.53)	(0.53)	(0.51)
Log(Husband income)	-3.08	-2.90		-3.07	-2.90	-2.98	-2.85	-3.29	-3.16
	(0.69)	(0.67)		(0.69)	(0.67)	(0.70)	(0.69)	(0.69)	(0.68)

Notes: All the specifications include time dummies and a quadratic function of age. FE denotes uncorrected fixed effects estimator; BC3 denotes the bias-corrected estimator proposed in the paper; MML denotes Carro (2003) Modified Maximum Likelihood estimator; LPM - FE denotes adjusted linear probability model (see text); LPM-BC denotes Nickell (1981) bias-corrected adjusted linear probability model; LPM-FE-FS denotes linear probability model; LPM-BC-FS denotes Nickell (1981) bias-corrected linear probability model. Column [3] taken from Carro (2003). The specification in Carro (2003) includes also a lag of Kids 0-2 with estimated coefficient -0.039 (0.054). Logit estimates and standard errors of index coefficients are normalized to have the same scale as probit. First period used as initial condition.
Source: PSID 1979-1988.