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

Iván Fernández-Val
Department of Economics
Boston University

February 4, 2009

Abstract

Fixed effects estimators of nonlinear panel models can be severely biased due to the incidental parameters problem. In this paper, I characterize the leading term of a large-T expansion of the bias of the MLE and estimators of average marginal effects in parametric fixed effects panel binary choice models. For probit index coefficients, the former term is proportional to the true value of the coefficients being estimated. This result allows me to derive a lower bound for the bias of the MLE. I then show that the resulting fixed effects estimates of ratios of coefficients and average marginal effects exhibit no bias in the absence of heterogeneity and negligible bias for a wide variety of distributions of regressors and individual effects in the presence of heterogeneity. I subsequently propose new bias-corrected estimators of index coefficients and marginal effects with improved finite sample properties for linear and nonlinear models with predetermined regressors.

JEL Classification: C23; C25; J22.

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

Acknowledgment. This is a revised version of the first chapter of my MIT PhD dissertation. I wish to thank Josh Angrist, Victor Chernozhukov, and Whitney Newey for encouragement and advice. I am grateful to Jesús Carro for sharing his data. For insightful suggestions and comments, I thank Paula Bustos, Sylvain Chassang, Michael Jansson, Guido Kuersteiner, Arthur Lewbel, Adam McCloskey, Frank Vella, three anonymous referees, an Associate Editor, and the Editor Cheng Hsiao. I have benefited from comments from seminar participants at Banco de España, Boston University, Brown University, CEMFI, Chicago GSB, Harvard KSG, LSE, MIT, Princeton University, UAB, UBC, Universidad de Navarra, University of Chicago, UPF, 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 unobserved individual characteristics are replaced by sample estimates, biasing estimates of model parameters. Examples include estimators for probit models 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; Katz, 2001; Greene, 2004; and Hahn and Newey, 2004).

In this paper, I develop bias corrections for fixed effects conditional maximum likelihood estimators (MLEs) in parametric panel binary choice models. These corrections are based on expressions for the bias that intensively exploit the structure of the problem by taking expectations using the conditional parametric model. Observed quantities are therefore replaced by expected quantities in the estimation of bias. This approach is similar to the use of the conditional information matrix in the estimation of the asymptotic variances of MLEs, instead of other alternatives such as the sample average of the outer product of the scores or the sample average of the negative Hessian (see Porter, 2002). Numerical results show that this refinement improves the finite sample performance of the correction over other existing methods.

In the case of probit index coefficients, I find that the bias of the conditional MLE is the product of a matrix and the true value of these coefficients, plus a second order term. This result allows me to derive a lower bound for the first order bias, depending uniquely upon the number of time periods in the panel. This bound establishes, for example, that the bias is at least 20% for 4-period panels and 10% for 8-period panels. When there is a single regressor, the above product matrix is a positive scalar and the probit fixed effects estimates are therefore biased away from zero, providing a theoretical foundation for previous numerical evidence (see, for e.g., Greene, 2004). In the case of no individual heterogeneity, the product matrix is a positive scalar multiple of the identity matrix, suggesting that the fixed effects estimators of ratios of coefficients do not suffer from incidental parameters bias when the level of heterogeneity is moderate. These ratios are structural parameters of interest because they can be interpreted as marginal rates of substitution in economic applications for which the index coefficients are only identified up to scale.

In nonlinear models one must often go beyond estimation of model parameters to 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 outcome. Accordingly, I also study the properties of fixed effects estimators of average marginal effects.[1] The motivation for my 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 (2002) conjectures that the estimators of the marginal effects have reasonable properties. Here, I characterize the analytical expression of the leading term of a large-T expansion of the bias of these average marginal effects. As Wooldridge anticipated, this bias is negligible relative to the true average marginal 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) (HN04 henceforth) find in Monte Carlo examples.

Most of the theoretical results I derive in this 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 of index coefficients and average marginal effects are biased downward for lagged dependent variables. This finding for marginal effects in dynamic nonlinear models resembles the analogous result for fixed effects estimators of model parameters in dynamic linear models. I subsequently develop new bias correction methods for estimates of index coefficients and marginal effects in probit and logit dynamic models that exhibit better finite sample properties than the existing alternatives. 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 by an analysis of female labor force participation, using 10 waves from the Panel Survey of Income Dynamics (PSID). This analysis is motivated by similar studies in labor economics in which panel binary choice processes have been widely used to model female labor

[1]Marginal effects are defined either as the change in the conditional outcome probability as a response to an one-unit increase in a regressor or as a local approximation to this quantity based on the slope of the conditional outcome probability. For example, in a probit model with a single regressor and constant term, the marginal effect can be defined either as $\Phi(\alpha + (x+1)\theta) - \Phi(\alpha + x\theta)$ or $\theta\phi(\alpha + x\theta)$, 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).

force participation decisions (see, e.g., Hyslop, 1999; Chay and Hyslop, 2000; and Carro, 2007). I find that fixed effects estimators, while biased for index coefficients, give very similar estimates to their bias-corrected counterparts of marginal effects in static models. On the other hand, uncorrected fixed effects estimators of both index coefficients and marginal effects are biased in dynamic models that account for true state dependence. In this case, the bias corrections I present in this paper are effective in 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 including, e.g., Phillips and Moon (1999), Lancaster (2002), Hahn and Kuersteiner (2002), Woutersen (2002), Arellano (2003), Alvarez and Arellano (2003), Hahn and Kuersteiner (2004), HN04, and Carro (2007); see also Arellano and Hahn (2007) for a recent survey of this literature and additional references. 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.[3] The idea behind these methods is to expand the incidental parameters bias of the estimator on the order of magnitude T, and to subtract an estimate of the leading term of the bias from the estimator.[4] As a result, the adjusted estimator has a bias of order T^{-2}, whereas the order of bias of the initial estimator is T^{-1}. All the above papers focus mainly on estimation of model parameters. The contribution of this paper is to provide a refinement of the bias corrections for parametric binary choice models with improved finite-sample properties and to develop theoretical results for the bias of index coefficients and marginal effects in probit models that help explain previous numerical findings.

The paper is organized as follows. Section 2 describes the problem of fixed effects estimation in panel binary choice models and develops bias corrections for these models. Section 3 looks at fixed effects estimation of marginal effects and establishes the small bias property for the probit model. Monte Carlo results and an empirical illustration are given in Sections 4 and 5, respectively. Section 6 concludes. The proofs of the main results are given in the appendix.

[3] Fixed-T-consistent estimators have also been derived for panel logit models (see Cox, 1958, Rasch, 1960, Andersen, 1973, and 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, thus precluding estimation of other quantities of interest such as marginal effects.

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

2 Fixed Effects Estimation of Panel Binary Choice Models

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:

$$Y_{it} = \mathbf{1}\left\{X_{it}'\theta_0 + \alpha_i - \epsilon_{it} \geq 0\right\}, \text{ for } i = 1, ..., n \text{ and } \ t = 1, ..., T,$$

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, and fertility (see Amemiya, 1981, for a survey).

2.2 Fixed Effects Estimators

In economic applications, regressors and individual heterogeneity are usually correlated because regressors typically include 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 realizations of the individual effects $\{\alpha_i\}_{i=1,...,n}$ as parameters to be estimated; see Mundlak (1978) for a similar interpretation of fixed effects estimators. Thus, assuming that the 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

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

where $F_{it}(\theta, \alpha_i)$ denotes $F_\epsilon(X_{it}'\theta + \alpha_i|X_i^t, \alpha_i)$.[5] The (conditional) MLE of θ, concentrating out the α_i's, is the solution to[6]

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

[5]In what follows, for any random variable Z, Z_{it} denotes the observation at period t for individual i and Z_i^t denotes a vector with all the observations for individual i up through period t, i.e., $Z_i^t := \{Z_{i1}, ..., Z_{it}\}$.

[6]Since inference is conditional on realizations of the regressors and individual effects, all probability statements should be qualified with almost surely. I omit this qualifier for notational convenience.

I only assume that the model is dynamically complete and that the regressor vector X is predetermined, rather than (strictly) exogenous. This is an important departure from the usual nonlinear panel data modeling assumptions as it permits, for instance, the model to capture richer dynamic feedbacks from the dependent variable to the regressors. The leading, but not exclusive, cases of predetermined variables are lagged dependent variables. Arellano and Carrasco (2003), for example, argue against the plausibility of the exogeneity assumption for fertility variables on the female labor participation decision and propose a random effects estimator that only requires the regressors to be predetermined. Wooldridge (2001) derives a fixed-T consistent conditional maximum likelihood estimator for panel models with predetermined regressors by imposing a parametric assumption on the conditional distribution of individual heterogeneity given the initial values of the predetermined regressors. Honoré and Lewbel (2002) and Lewbel (2005) propose alternative fixed-T consistent estimators for panel binary choice models with predetermined regressors in the case for which one of the regressors is exogenous and independent of the individual effect (conditional on the other regressors). Carro (2007) proposes a large-T bias corrected estimator for dynamic binary choice models that permits lags of the dependent variable to be regressors but assumes that the other explanatory variables are exogenous. Here, I propose large-T bias corrected fixed effects estimators for static and dynamic binary choice models with general predetermined regressors without imposing parametric assumptions on individual heterogeneity.

2.3 The 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 individual effects. Hence, the estimation error of the individual effects contaminates the estimates of the 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,

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

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 in this paper, $\theta_T = \theta_0 + \mathcal{B}/T + O(T^{-2})$ for some $\mathcal{B}$ (see HN04 for details). By the asymptotic normality of the MLE, $\sqrt{nT}(\widehat{\theta} - \theta_T) \xrightarrow{d} N(0, \mathcal{J}^{-1})$ as $n \to \infty$, where $\mathcal{J}$ is defined below. Therefore,

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

Here we can see that even if we let T grow at the same rate as n, the MLE, while consistent, has a limiting distribution not centered around the true parameter value. Intuitively, under these asymptotic sequences, the estimates of the individual effects converge to their true values at rate $\sqrt{T}$ slower than $\sqrt{nT}$, 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 a consequence of the large-T version of the incidental parameters problem, which invalidates standard inference based on this distribution.

2.4 Large-T Approximation of the Bias

Hahn and Kuersteiner (2004) (HK04 henceforth) characterize the expression of the bias $\mathcal{B}$ using a stochastic expansion of the fixed effects estimator on the order of T for general dynamic models. Here, I briefly reproduce this expansion and the resulting expression for the bias to introduce notation that will be extensively used in the subsequent analysis. Let

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

and additional subscripts denote partial derivatives, e.g., $u_{it\theta}(\theta, \alpha) := \partial u_{it}(\theta, \alpha)/\partial\theta'$. 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 (leading term of the large-T expansion of the) asymptotic bias is

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

where $\mathcal{J} := E_n[\mathcal{J}_i]$ is the probability limit of minus the Jacobian of the estimating equation for θ, i.e.,

$$\mathcal{J}_i := -\left\{E_T\left[u_{it\theta}\right] - E_T\left[u_{it\alpha}\right]E_T\left[v_{it\theta}\right]/E_T\left[v_{it\alpha}\right]\right\}, \tag{2.1}$$

with $E_T\left[h_{it}\right] := \lim_{T\to\infty}\sum_{t=1}^{T} h_{it}/T$, and $b := E_n[b_i]$ is the bias of the estimating equation for θ, i.e.,

$$b_i := E_T\left[u_{it\alpha}\right]\beta_i + \bar{E}_T\left[u_{it\alpha}\psi_{is}\right] + E_T\left[u_{it\alpha\alpha}\right]\sigma_i^2/2, \tag{2.2}$$

with $\bar{E}_T[h_{it}k_{is}] := \sum_{j=-\infty}^{\infty} E_T[h_{it}k_{i,t-j}]$, the spectral expectation.[7] Here, σ_i^2 and β_i are the asymptotic variance and bias components of a higher order expansion for the estimator of the individual effects. That is, as $T \to \infty$,

$$\begin{aligned} \widehat{\alpha}_i &= \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + o_p(1/T), \ \psi_i = \sum_{t=1}^{T} \psi_{it}/\sqrt{T} \xrightarrow{d} \mathcal{N}(0, \sigma_i^2), \\ \psi_{it} &= -E_T[v_{it\alpha}]^{-1} v_{it}, \quad \sigma_i^2 = \bar{E}_T[\psi_{it}\psi_{is}], \quad \text{and} \\ \beta_i &= -E_T[v_{it\alpha}]^{-1} \left\{ \bar{E}_T[v_{it\alpha}\psi_{is}] + E_T[v_{it\alpha\alpha}]\sigma_i^2/2 \right\}. \end{aligned} \tag{2.3}$$

2.5 Bias Corrections for Panel Binary Choice Models

Large-T correction methods reduce the order of the bias from $O(T^{-1})$ to $O(T^{-2})$ by removing an estimate of $\mathcal{B}/T$ from the fixed effects estimator. These corrections center the asymptotic distribution at the true parameter value under $T/n^{1/3} \to \infty$. To see this, note that if $\widehat{\theta}^c$ is a bias corrected estimator of θ_0 with probability limit $\theta_T^c = \theta_0 + O\left(T^{-2}\right)$, then

$$\sqrt{nT}(\widehat{\theta}^c - \theta_0) = \sqrt{nT}(\widehat{\theta}^c - \theta_T^c) + O_p\left(\sqrt{n/T^3}\right).$$

These methods can take different forms depending on whether the adjustment is made in the estimator (as in HK04 and HN04), estimating equation (as in Woutersen, 2002, and Fernández-Val, 2004), or objective function (as in Arellano and Hahn, 2006). Alternative panel Jackknife corrections, which do not require explicit characterization of the bias expression, have been proposed by HN04 for static models and Dhaene, Jochmans, and Thuysbaert (2006) for dynamic models.

Expressions for the bias can be further specialized in parametric binary choice models. In these models, the scores of the conditional log-likelihood for observation i at time t take the forms

$$v_{it}(\theta, \alpha_i) = H_{it}(\theta, \alpha_i)\left(Y_{it} - F_{it}(\theta, \alpha_i)\right) \quad \text{and} \quad u_{it}(\theta, \alpha_i) = v_{it}(\theta, \alpha_i) X_{it},$$

where $H_{it}(\theta, \alpha_i) = f_{it}(\theta, \alpha_i)/\left[F_{it}(\theta, \alpha_i)\left(1 - F_{it}(\theta, \alpha_i)\right)\right]$, $f_{it}(\theta, \alpha_i)$ denotes $f_\epsilon(X_{it}'\theta + \alpha_i | X_i^t, \alpha_i)$, and f_ϵ is the pdf associated with F_ϵ. Moreover, by the Law of Iterated Expectations, $E_T[h(Z_{it})] = E_T\left[E\left[h(Z_{it})|X_i^t, \alpha\right]\right]$ for any function $h(Z_{it})$. Taking conditional expectations of the expressions

[7] The previous expansion provides an alternative explanation for the absence of incidental parameters bias in the linear panel model with exogenous regressors and $Y_{it}|X_i^T, \alpha_i \sim i.i.d.\mathcal{N}(X_{it}'\theta_0 + \alpha_i, \sigma_\epsilon^2)$. In this case, $v_{it} = (Y_{it} - X_{it}'\theta_0 - \alpha_i)/\sigma_\epsilon^2$, $v_{it\alpha} = -1/\sigma_\epsilon^2$, $v_{it\alpha\alpha} = 0$, $u_{it} = v_{it}X_{it}$, $u_{it\alpha} = v_{it\alpha}X_{it}$, and $u_{it\alpha\alpha} = 0$. Then, $\beta_i = b = 0$ since $\bar{E}_T[v_{it\alpha}\psi_{is}] = \bar{E}_T[u_{it\alpha}\psi_{is}] = 0$ 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.

for the components of the bias in expressions (2.1)-(2.3) yields

$$\sigma_i^2 = E_T\left[H_{it} f_{it}\right]^{-1}, \quad \psi_{it} = H_{it}(Y_{it} - F_{it})\sigma_i^2, \tag{2.4}$$

$$\beta_i = -E_T\left[H_{it} g_{it}\right]\sigma_i^4/2 - \tilde{E}_T[H_{it} f_{it} \psi_{is}]\sigma_i^2, \tag{2.5}$$

$$b_i = -E_T\left[H_{it} f_{it} X_{it}\right]\beta_i - E_T\left[H_{it} g_{it} X_{it}\right]\sigma_i^2/2 - \tilde{E}_T\left[H_{it} f_{it} X_{it} \psi_{is}\right], \quad \text{and} \tag{2.6}$$

$$\mathcal{J}_i = E_T\left[H_{it} f_{it} X_{it} X_{it}'\right] - E_T\left[H_{it} f_{it} X_{it}\right] E_T\left[H_{it} f_{it} X_{it}'\right]\sigma_i^2, \tag{2.7}$$

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 $g_\epsilon(X_{it}'\theta + \alpha_i | X_i^t, \alpha_i)$, and g_ϵ is the derivative of f_ϵ. If the regressors are exogenous, the terms involving $\tilde{E}_T$ drop out. Note that these expressions differ from the bias expressions in HK04 and HN04 because they exploit the fact that in parametric binary choice models, the generalized residuals v_{it} and u_{it} follow martingale differences. Accordingly, terms involving covariances of these residuals with past and present values of functions of the regressors and individual effects are zero.

To describe how to construct the correction from the new bias formulas, it is convenient to introduce some more notation. Let $\widehat{F}_{it}(\theta) := F_{it}(\theta, \widehat{\alpha}_i(\theta))$, $\widehat{f}_{it}(\theta) := f_{it}(\theta, \widehat{\alpha}_i(\theta))$, $\widehat{g}_{it}(\theta) := g_{it}(\theta, \widehat{\alpha}_i(\theta))$, and $\widehat{H}_{it}(\theta) := H_{it}(\theta, \widehat{\alpha}_i(\theta))$. Also, define

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

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 in (2.4). Let

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

$$\widehat{\mathcal{J}}_i(\theta) := \widehat{E}_T\left[\widehat{H}_{it}(\theta)\widehat{f}_{it}(\theta)X_{it}X_{it}'\right] - \widehat{E}_T\left[\widehat{H}_{it}(\theta)\widehat{f}_{it}(\theta)X_{it}\right]\widehat{E}_T\left[\widehat{H}_{it}(\theta)\widehat{f}_{it}(\theta)X_{it}'\right]\widehat{\sigma}_i^2(\theta),$$

where $\widehat{\tilde{E}}_{T,J}[h_{it}k_{is}] := \sum_{j=1}^{J}\sum_{t=j+1}^{T} h_{it}k_{i,t-j}/(T-j)$ and J is a bandwidth parameter for spectral expectation estimation that needs to be chosen such that $J/T^{1/2} \to 0$ as $T \to \infty$; see Hahn and Kuersteiner (2007) for details on bandwidth choice in the context of bias estimation. Here, $\widehat{\beta}_i(\theta)$ is an estimator of the higher-order asymptotic bias of $\widehat{\alpha}_i(\theta)$ in (2.5) and $\widehat{\mathcal{J}}(\theta) := \sum_{i=1}^{n}\widehat{\mathcal{J}}_i(\theta)/n$ is an estimator of minus the Jacobian of the estimating equation for θ in (2.7). Then, the estimator of $\mathcal{B}$ is

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

where $\widehat{b}(\theta) := \sum_{i=1}^{n}\widehat{b}_i(\theta)/n$ and

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

is an estimator of the bias of the estimating equation for θ in (2.6).

The one-step bias corrected estimator can then be formed by evaluating the previous expression at the MLE, that is, $\widehat{\mathcal{B}} = \widehat{\mathcal{B}}(\widehat{\theta})$ and $\tilde{\theta} = \widehat{\theta} - \widehat{\mathcal{B}}/T$. The 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 also be obtained by solving the modified first order condition:

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

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

2.6 Bias for the Panel Probit Model

The expression for the bias takes a simple form for the panel probit model with exogenous regressors, helping explain the results of previous numerical studies (Greene, 2004; and HN04). In particular, this bias is the product of a matrix and the true value of the parameter plus a second order term. This product matrix can be characterized in some special situations. For example, when there is a single regressor (as in the studies aforementioned) the matrix is a positive scalar and the fixed effects estimator is therefore biased away from zero. In the absence of heterogeneity, this property also holds, regardless of the dimension of the regressor vector, because the product matrix is a positive scalar multiple of the identity matrix.

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) $E_T[W_{it}W_{it}']$ exists and is nonsingular for each i with $W_{it} = (1, X_{it}')$; (iii) $\{X_{it}\}_{t=1}^{T}$ is stationary and strongly mixing with mixing coefficients $a_i(m)$ such that $\sup_i |a_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$; (iv) $\{Y_i^T, X_i^T, \alpha_i\}_{i=1}^{n}$ are independent; (v) $\sup_i E_T\left[|X_{it}^j|^{42+5p+\delta}\right] < \infty$ for $j \in \{1, ..., p\}$ and some $\delta > 0$, where X_{it}^j is the j-th element of X_{it}; (vi) $n = o(T^3)$; and (vii) $\{X_i^T\}_{i=1}^{n}$ is identically distributed conditional on α_i.*[8] *Then,*

1.
$$\mathcal{B} = E_n[\mathcal{J}_i]^{-1} E_n\left[\mathcal{J}_i \sigma_i^2\right] \theta_0/2 \quad \textit{and} \tag{2.8}$$

2. $\alpha_i = \alpha \; \forall i$ *implies*
$$\mathcal{B} = \sigma^2 \theta_0/2,$$

[8] A sequence $\{X_t\}$ is strongly mixing with mixing coefficients $a(m)$ if $\sup_t \sup_{A \in \mathcal{A}_t, E \in \mathcal{E}_{t+m}} |P(A \cap E) - P(A)P(E)| = a(m)$, where $\mathcal{A}_t = \sigma(X_t, X_{t-1}, ...)$, $\mathcal{E}_t = \sigma(X_t, X_{t+1}, ...)$, and $\sigma(X_t, ..., X_s)$ denotes the sigma-algebra generated by $\{X_t, ..., X_s\}$.

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. ■

Condition (i) is the probit modeling assumption and imposes exogeneity of the regressors; condition (ii) is standard for MLEs (Newey and McFadden, 1994) and guarantees identification and asymptotic normality of MLEs of model parameters and individual effects based on time series variation; assumption (iii), together with (iv) and the moment condition (v), is imposed in order to apply a law of large numbers to the cross sections; and assumptions (v) and (vi) guarantee the validity of the higher order expansions for the fixed effects estimators (see, e.g., Example 1 in HK04). Condition (vii) is imposed to guarantee that the regressors follow the same process for all the individuals in the absence of heterogeneity.

When there is a single regressor, σ_i^2 and $\mathcal{J}_i$ are positive scalars and the bias is a positive multiple of the parameter value. Moreover, since $\sigma_i^2 \geq \Phi(0)\left[1-\Phi(0)\right]/\phi(0)^2 = \pi/2$, the bias $\mathcal{B}$ is bounded from below when there is a single regressor or no heterogeneity.

Corollary 1 *Under the conditions of Proposition 1, if* $p = 1$ *or* $\alpha_i = \alpha\ \forall i$*, then for any* $j \in \{1,...,p\}$ *with* $\theta_0 = (\theta_{0,1},...,\theta_{0,p})'$ *and* $\mathcal{B} = (\mathcal{B}_1,...,\mathcal{B}_p)'$*,*

$$|\mathcal{B}_j| \geq \pi\ |\theta_{0,j}|/4.$$

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.

Table 1 compares the large-n asymptotic biases of fixed effects estimators to their first order approximation based on expression (2.8) of Proposition 1 for panels with 4 and 8 time periods. I consider a model with a single regressor, $\theta_0 = 1$, and an independent individual effect with several distributions. For the regressor, the distributions are normalized to have the same mean and variance as the Nerlove process, defined in (4.1); mean 0.35 and variance 0.15 for $T = 4$, and mean 0.72 and variance 0.30 for $T = 8$. For the individual effect, the distributions are normalized to have zero mean and unit variance. However, $C = 0$ denotes a degenerate distribution at zero, corresponding to the absence of heterogeneity ($\alpha_i = \alpha\ \forall i$). All the entries of the table are expressed in percentage of the true parameter value. The asymptotic biases lie between 33 and 41% for $T = 4$, and between 13 and 21% for $T = 8$. The first order approximation captures between 62 and 74% of the bias for $T = 4$, and between 75 and 100% of the bias for $T = 8$.

The second expression of the above 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 linear panel 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 model 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

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

where the last equality follows from a standard first order Taylor expansion of $(1-1/T)^{-1/2}$ around $1/T = 0$. Here we can see a parallel with the probit model, 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 cannot be separated from the estimation of β_0 in the probit case. In other words, over-fitting due to the fixed effects estimation biases the estimates of the model parameter upward because the standard deviation is implicitly in the denominator of the model parameter estimated by the probit regression.

Proportionality implies, in turn, zero bias for fixed effects estimators of ratios of index coefficients. These ratios can be behavioral parameters of interest because they are direct measures of the relative effects of the regressors and can be interpreted as marginal rates of substitution in economic applications (see, e.g., McFadden, 1974). In general though, 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. The bias for the ratio of index coefficients shrinks to zero as the variance of the underlying distribution of individual effects decreases toward zero.

3 Fixed Effects Estimation of Marginal Effects

3.1 Marginal Effects

In empirical analysis, the ultimate quantities of interest are often the marginal effects of specific changes in the regressors on the conditional response probability (see, e.g., Ruud, 2000; Angrist, 2001; Wooldridge, 2002; and Wooldridge, 2005). Index coefficients in binary choice models provide information about the signs and relative magnitudes of these effects, but not their absolute magnitudes. Thus, for example, the marginal effect of the first regressor X_1 on the conditional probability that $Y = 1$ for individual i at time t, evaluated at $x_{it} = (x_{1it}, x_{2it}')$, is given by

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

where $\theta = (\theta_1, \theta_2')'$. When X_1 is discrete, the marginal effect is usually calculated as the effect of a one-unit increase in the regressor on the conditional probability:

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

Marginal effects in nonlinear models depend on the individual effects α_i and the level chosen for evaluating the regressors x_{it}. This heterogeneity raises the question of what the relevant effects to report are. A common practice is to give some summary measure, such as the average effect or the effect for some interesting values of the regressors. Chamberlain (1984) suggests reporting the average effect for an individual randomly drawn from the population. That is,

$$\mu(\theta) = \int m(X_{it}, \theta, \alpha_i) dH_{X_{it}, \alpha_i}(X_{it}, \alpha_i) \tag{3.1}$$

or

$$\mu(x_1, \theta) = \int \tilde{m}((x_1, X_{2it}')', \theta, \alpha_i) dG_{X_{2it}, \alpha_i}(X_{2it}, \alpha_i), \tag{3.2}$$

where H and G are the joint distributions of (X_{it}, α_i) and (X_{2it}, α_i), respectively, and x_1 is some interesting value of X_1. These effects can also be calculated for subpopulations of interest by conditioning on relevant values of the covariates. For example, if X_1 is binary (e.g., a treatment indicator), the average treatment effect on the treated (ATT) is

$$\mu_1(\theta) = \int \tilde{m}((0, X_{2it}')', \theta, \alpha_i) dG_{X_{2it}, \alpha_i}(X_{2it}, \alpha_i | X_{1it} = 1). \tag{3.3}$$

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 implementation issues. For example, replacing population expectations by sample analogs does not always work in binary choice panel 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 whose response does not change status in the sample. Therefore, the sample average of the estimated individual effects is generally not well-defined. It should be noted here that the definitions of marginal effects given by (3.2), (3.1), and (3.3) are static in the sense that they do not account for possible feedback from present changes in the response variable to future values of the regressors. 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).

3.2 Bias Corrections for Marginal Effects

Fixed effects estimators of average marginal effects can be constructed from equations (3.1), (3.2), and (3.3) by replacing population moments by sample analogs and using fixed effects estimators of the individual effects. That is,

$$\widehat{\mu}(\theta) = \sum_{i=1}^{n} \widehat{E}_T \left[m \left(X_{it}, \theta, \widehat{\alpha}_i(\theta) \right) \right] / n, \tag{3.4}$$

$$\widehat{\mu}(x_1, \theta) = \sum_{i=1}^{n} \widehat{E}_T \left[\tilde{m} \left((x_1, X_{2it}')', \theta, \widehat{\alpha}_i(\theta) \right) \right] / n, \tag{3.5}$$

and

$$\widehat{\mu}_1(\theta) = \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{m} \left((0, X_{2it}')', \theta, \widehat{\alpha}_i(\theta) \right) \mathbf{1} \left\{ X_{1it} = 1 \right\} / N_1,$$

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

These fixed effects estimators generally suffer from the incidental parameters problem even when evaluated at the true value of the index coefficients θ_0. The source of this problem is the dependence of the estimators on the estimates of the individual effects $\widehat{\alpha}_i(\theta)$. As for model parameters estimates, 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 (2.3), we have, as $n \to \infty$ and $T \to \infty$,

$$T \left(\sum_{i=1}^{n} \widehat{E}_T \left[m \left(X_{it}, \theta_0, \widehat{\alpha}_i(\theta_0) \right) \right] / n - E_n E_T \left[m \left(X_{it}, \theta_0, \alpha_i \right) \right] \right) \xrightarrow{p} E_n \left[\Delta_i \right],$$

where

$$\Delta_i = E_T \left[m_\alpha \left(X_{it}, \theta_0, \alpha_i \right) \right] \beta_i + \tilde{E}_T \left[m_\alpha \left(X_{it}, \theta_0, \alpha_i \right) \psi_{is} \right] + E_T \left[m_{\alpha\alpha} \left(X_{it}, \theta_0, \alpha_i \right) \right] \sigma_i^2 / 2 \tag{3.6}$$

and subscripts on m denote partial derivatives.[9] 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 HN04.

To describe the bias corrections for marginal effects, let $\tilde{\theta}$ be a bias-corrected estimator of θ_0 and $\tilde{\alpha}_i = \widehat{\alpha}_i(\tilde{\theta})$, $i = 1, ..., n$, be the corresponding estimators of the individual effects. Then, large-T consistent estimators for the individual components of the bias in equation (3.6) can be formed by replacing population moments by sample analogs and true parameter values by bias corrected estimates. That is,

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

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

The bias corrected estimator for μ can then be formed as

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

3.3 Panel Probit: Small Bias Property

HN04 find small biases in marginal effects constructed from uncorrected fixed effects estimators of index coefficients for the probit model in numerical examples. The following proposition characterizes the analytical expression for the leading term of the bias for uncorrected probit fixed effects estimators of average marginal effects in a model with exogenous regressors. This expression helps explain the previous numerical findings.

Proposition 2 (Bias for Marginal Effects) *Let* $\widehat{\mu} := \widehat{\theta} \ \sum_{i=1}^{n} \widehat{E}_T \left[\phi\left(X_{it}'\widehat{\theta} + \widehat{\alpha}_i(\widehat{\theta})\right)\right]/n$ *and* $\mu_0 = \theta_0 \ E_n \left\{E_T\left[\phi\left(X_{it}'\theta_0 + \alpha_i\right)\right]\right\}$*, then under the conditions of Proposition 1, as* $n, T \to \infty$*,*

$$\widehat{\mu} = \mu_0 + \mathcal{B}_\mu / T + O_p(T^{-2}),$$

where $\mathcal{B}_\mu = E_n\left[\omega_i\left(\mathcal{B}_i - \mathcal{B}\right)\right], \mathcal{B}_i = \sigma_i^2 \theta_0/2,$ *and* $\omega_i = E_T\left[\phi_{it}\left(\xi_{it}\theta_0\left(X_{it} - \sigma_i^2 E_T\left[H_{it}\phi_{it}X_{it}\right]\right)' - \mathcal{I}_p\right)\right]$, *with* $\xi_{it} = X_{it}'\theta_0 + \alpha_i$*,* $\phi_{it} = \phi(\xi_{it})$*, and* $\mathcal{I}_p$ *denotes a* $p \times p$ *identity matrix.*

Proof. See Appendix. ■

The key term in the expression of the bias is $(\mathcal{B}_i - \mathcal{B})$. This term captures deviations of the individual contributions to the bias $\mathcal{B}_i$, corresponding to the bias if all the individuals have value α_i for the individual effect and the same process for X_i^T, from a matrix weighted average of these contributions, corresponding to the bias $\mathcal{B} = E_n[\mathcal{J}_i]^{-1} E_n[\mathcal{J}_i \mathcal{B}_i]$.[10] The additional factor ω_i reduces the impact of individual contributions that are far from the weighted average. This is illustrated in Figure 1, which plots the components of the bias for independent standard normally distributed regressors and individual effects. The components ω_i and $(\mathcal{B}_i - \mathcal{B})$ have U shapes and the term ϕ_{it} in ω_i acts to reduce the weights in the tails of $(\mathcal{B}_i - \mathcal{B})$, where $(\mathcal{B}_i - \mathcal{B})$ is large. In this case, the bias $\omega_i(\mathcal{B}_i - \mathcal{B})$, as a function of α_i, is positive at zero and takes negative values as α_i moves away from the origin. The positive and negative values of $\omega_i(\mathcal{B}_i - \mathcal{B})$ offset each other when they are integrated, using the distribution of the individual effects, to obtain $\mathcal{B}_\mu$.

In the absence of heterogeneity, the $\mathcal{B}_i$'s are constant and the bias $\mathcal{B}_\mu$ is thus zero. This observation indicates that, as in linear models, the inclusion of irrelevant variables (fixed effects),

[10] The expression for $\mathcal{J}_i$ for the probit model is given in the proof of Proposition 1 in the Appendix.

while reducing efficiency, does not affect the consistency of the probit estimates of marginal effects. In general, the bias of the estimator of the average marginal effect depends upon the degree of variability of the individual effects and the heterogeneity in the distributions of regressors across individuals. These factors determine the degree of heterogeneity in the $\mathcal{B}_i$'s and $\mathcal{J}_i$'s, the weights for each individual in the estimator of the index coefficient θ_0.

Table 2 compares the large-n asymptotic biases for marginal effects to their first order approximation based on the expression of Proposition 2 for panels with 4 and 8 time periods. The parameters and distributions considered for the regressors and individual effects are the same as those in Table 1. All the entries of the table are expressed as a percentage of the true marginal effect. Here, the bias for the marginal effects estimator is never higher than 1% for both $T = 4$ and $T = 8$. The first order approximations are relatively less accurate in percentage terms than those for index coefficients but the biases are very small in this case.

The small bias property for fixed effects estimators of marginal effects does not carry over to dynamic models or, more generally, to models with predetermined regressors. The reason is that, due to dynamic feedbacks, averaging across individuals does not fully remove the additional bias components. To understand this result, we can look at the survivor probabilities at zero in a dynamic Gaussian linear model. The analysis here is motivated by HN04, which finds no asymptotic bias in the fixed effects estimation of these probabilities in a static Gaussian linear model. Specifically, suppose that $Y_{it} = \theta_0 Y_{i,t-1} + \alpha_i + \epsilon_{it}$, where $\epsilon_{it}|Y_{i,t-1}, ..., Y_{i,0}, \alpha_i \sim N(0, \sigma_\epsilon^2)$, $Y_{i0}|\alpha_i \sim \mathcal{N}(\alpha_i/(1-\theta_0), \sigma_\epsilon^2/(1-\theta_0^2))$, and $0 \leq \theta_0 < 1$. The survivor probability evaluated at $Y_{i,t-1} = r$ and its fixed effects estimator are

$$S_0 = E_n\left\{\Phi\left(\frac{\theta_0\ r + \alpha_i}{\sigma_\epsilon}\right)\right\} \quad \text{and } \widehat{S} = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{\widehat{\theta}\ r + \widehat{\alpha}_i(\widehat{\theta})}{\widehat{\sigma}_\epsilon}\right),$$

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

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

Then, as $n \to \infty$

$$
\begin{aligned}
\widehat{S} &\xrightarrow{p} E_n \left\{ E \left[\Phi \left(\frac{\theta_T \ r + \widehat{\alpha}_i(\theta_T)}{\sigma_{\epsilon T}} \right) \right] \right\} \\
&= E_n \left\{ E \left[\Phi \left(\frac{\theta_0 r + \alpha_i + v_i + (\theta_T - \theta_0) \left(r - \frac{\alpha_i}{1-\theta_0} \right) + o_p(T^{-1})}{\sigma_{\epsilon T}} \right) \right] \right\} \\
&= E_n \left\{ \Phi \left[\frac{\theta_0 r + \alpha_i}{\sigma_\epsilon} - \frac{1}{T} \frac{1+\theta_0}{\sigma_\epsilon} \left(r - \frac{\alpha_i}{1-\theta_0} \right) \right] + o_p(T^{-1}) \right\}, \qquad (3.7)
\end{aligned}
$$

by the convolution properties of the normal distribution (see Lemma 1 in Appendix A).

In expression (3.7) we can see that averaging across individuals eliminates the asymptotic bias from $\widehat{\sigma}_\epsilon^2$, but does not affect the bias from $\widehat{\theta}$. The sign of the bias of $\widehat{S}$ generally depends on the evaluation value r and the distribution of the individual effects. When Y_{it} is binary, for example, $\widehat{S}$ underestimates (overestimates) the underlying survivor probability when evaluated at values above (below) the unconditional means of the response, $\alpha_i/(1-\theta_0)$. This implies that if marginal effects are computed as differences in survivor probabilities evaluated at two different values (see, e.g., expression (3.2) for discrete regressors), the fixed effects estimates of marginal effects will be biased downward if the values chosen are $r = 0$ and $r = 1$.[11] For exogenous variables, Monte Carlo results in Section 4 suggest that the bias problem is less severe. Intuitively, it seems that the part of the bias due to dynamic feedbacks is less important for this type of regressor.

4 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 estimators is illustrated for several lengths of the panel. All the results presented here are based on 1000 replications and the designs follow Heckman (1981), Greene (2004), and HN04 for the static probit model, and Honoré and Kyriazidou (2000), HK04, and Carro (2007) for the dynamic logit model.[12]

[11] Note that when Y_{it} is binary, all the unconditional means are between zero and one.

[12] Heckman (1981) and Greene (2004) use $X_{i0} = 5 + 10u_{i0}$ as the initial condition for the Nerlove process in (4.1). I use the same initial condition as in HN04.

4.1 Static Probit Model

The model design is

$$\begin{aligned} Y_{it} &= \mathbf{1}\{X_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\}; \ \epsilon_{it} \sim i.i.d.\mathcal{N}(0,1); \ \alpha_i \sim i.i.d.\mathcal{N}(0,1); \\ X_{it} &= t/10 + X_{i,t-1}/2 + u_{it} \ \text{ for } t = 1, ..., T; \ X_{i0} = u_{i0}; \ \ u_{it} \sim i.i.d.\mathcal{U}(-1/2, 1/2); \\ n &= 100; \ T = 4, \ 8, \ 12; \ \theta_0 = 1; \end{aligned} \tag{4.1}$$

where $\mathcal{N}$ and $\mathcal{U}$ denote normal and uniform distributions, 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 the standard deviation of the estimators; and MAE denotes median absolute error.[13] $PROBIT$ corresponds to the uncorrected MLE, $BC1$ and $BC2$ correspond to the one-step analytical bias-corrected estimators of HN04 based on a maximum likelihood setting and general estimating equations, respectively. JK is the bias correction based on the leave-one-period-out Jackknife-type estimator, see HN04. $BC3$ is the one-step bias-corrected estimator proposed here. Iterated bias-corrected and score-corrected estimators are not considered because they are much more computationally cumbersome.[14]

Table 3 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, 2004 and HN04) which show that the probit MLE is severely biased, even when $T = 12$, inducing significant distortions in rejection probabilities. $BC3$ has substantially lower bias than all other estimators and improves rejection probabilities over HN04's analytical and jackknife bias-corrected estimators for small sample sizes.[15] All bias corrections reduce dispersion. This can be explained by the proportionality result for the bias in Proposition 1. To check the sensitivity of the bias corrections to the condition $n = o(T^3)$ used in the asymptotic analysis, I repeated the simulations for $n = 1,000$. The estimates of the biases are very similar to the ones reported here, for $n = 100$.[16]

---

[13] The mean of the MLE is generally not well-defined for finite n because complete data separation might occur with positive probability. In this case the MLE is either ∞ or $-\infty$. In the simulations I only find one of these cases for $T = 4$, which is discarded for the computation of the reported mean. This problem motivates to report the median and median absolute error, instead of root mean squared error, as an overall measure of goodness of fit. I thank a referee for pointing out this general issue with discrete choice models.

[14] HN04 find that iterating the bias correction does not matter much in this example.

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

[16] These results are not reported for brevity's sake and are available from the author upon request.

Table 4 reports the ratio of marginal effect estimators to their true value. The estimands correspond to the derivatives of the response probability averaged across individuals and time periods, as defined in Proposition 2. These effects are different from the marginal effects reported in HN04, which are the derivatives of the response probability evaluated at the individual means of the regressor and averaged across individuals. It can be shown that a similar small bias property to Proposition 2 holds for the marginal effects of HN04. Here, I also include two estimators of the average marginal effect based on linear probability models. $LPM - FS$ is the standard linear probability model that uses all the observations, and LPM is an adjusted version that calculates the slope from individuals that change status during the sample, i.e., including only individuals with $Y_{it} = 1$ for some t and $Y_{is} = 0$ for some s, and assigns zero effect to the rest of individuals in the average effect. The results show small biases in uncorrected fixed effects estimators of marginal effects. Rejection frequencies are higher than their nominal levels due to the underestimation of dispersion, apparent in the values of SE/SD. As in cross-section models (e.g., Angrist, 2001 and Hahn, 2001), both linear models work fairly well in estimating the average marginal effect.[17]

4.2 Dynamic Logit Model

The model design is

$$\begin{aligned} Y_{i0} &= \mathbf{1}\{\theta_{X,0}X_{i0} + \alpha_i - \epsilon_{i0} \geq 0\}; \\ Y_{it} &= \mathbf{1}\{\theta_{Y,0}Y_{i,t-1} + \theta_{X,0}X_{it} + \alpha_i - \epsilon_{it} \geq 0\}, \quad \text{for } t = 1, ..., T-1; \\ \epsilon_{it} &\sim i.i.d.\mathcal{L}(0, \pi^2/3); \quad X_{it} \sim i.i.d.\mathcal{N}\left(0, \pi^2/3\right); \\ n &= 250; \ T = 8, \ 12, \ 16; \ \theta_{Y,0} = .5; \ \theta_{X,0} = 1; \end{aligned}$$

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 in Section 4.1. $LOGIT$ is the uncorrected MLE; $BC1$ denotes the bias-corrected estimator of HK04; HK is the dynamic version of the conditional logit estimator of Honoré and Kyriazidou (2000), which is fixed-T consistent; MML is the Modified MLE for dynamic models of Carro (2007); and $BC3$ is the bias-corrected estimator that uses expected

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

quantities in the estimation of the bias formulas proposed in this paper.[18] For the number of lags, I choose the bandwidth parameter $J = 1$, as in HK04.

Tables 5 and 6 present the Monte Carlo results for the index coefficients $\theta_{Y,0}$ and $\theta_{X,0}$. Overall, all the bias-corrected estimators have substantitally smaller finite sample bias and much better inference properties than the uncorrected MLEs. Large-T-consistent estimators have median absolute errors comparable to HK for $T = 8$.[19] 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 more computationally attractive than MML as it does not require modification of the probit first order conditions.

Tables 7 and 8 report the Monte Carlo results for the estimators of average marginal effects of the lagged dependent variable and exogenous regressor, respectively. These effects are calculated using expression (3.5) with $x_1 = 0$ for the lagged dependent variable and expression (3.4) for the exogenous regressor. The results are expressed in percentage of the true parameter value. Here, I present results for the standard MLE Logit, $BC1$, $BC3$, 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 3, uncorrected estimates of the marginal effect 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 of the marginal effect of the lagged dependent variable.

5 Empirical Illustration: Female Labor Force Participation

The relationship between fertility and female labor force participation has been 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 at solving this omitted variables problem by controlling

[18] 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.

[19] An aspect not explored here is that the performance of HK estimator deteriorates with the number of exogenous variables. For example, HK04 find that their large-T-consistent estimator out-performs HK for $T = 8$ when the model includes two exogenous variables.

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 (2007).

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

$$P_{it} = \mathbf{1}\left\{\delta_t + P_{i,t-1}\theta_P + X_{it}'\theta_X + \alpha_i - \epsilon_{it} \geq 0\right\}, \tag{5.1}$$

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, the natural logarithm of the husband's earnings, and a quadratic function of age.[20]

The sample is selected from waves 13 to 22 of the Panel Study of Income Dynamics (PSID) and contains information for the ten calendar years 1979-1988. Only women aged 18-60 in 1985 who were continuously married with husbands in the labor force in each of the sample periods are included in the sample. The sample considered consists of 1,461 women, 664 of whom changed labor force participation status during the sample period. The first year of the sample is excluded for use as the initial condition in the dynamic model.

Descriptive statistics for the sample are given in Table 9. 21% of the sample is black and the average age in 1985 was 37. Roughly 72% of the women in the sample participated in the labor force at some period. The average years of schooling in the sample is 12 years. The average numbers of children per woman were .2, .3, and 1.1 for the three categories 0-2 year-old, 3-5 year-old, and 6-17 year-old, respectively.[21] On average, women that changed participation status during the sample, in addition to being younger, were less educated, less likely to be black, had more dependent children, and had higher spousal earnings. Interestingly, women who never participated did not have more children than women who were employed every year on average, 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 10 reports fixed effects estimates of index coefficients and marginal effects obtained from a static specification, that is, excluding the lag of participation in equation (5.1). Estima-

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

[21] Years of schooling were 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); and 8 = 'college and advanced degree' (18 years). See also Hyslop (1999).

tors are labeled as in the Monte Carlo example, with C denoting Andersen's (1973) conditional logit estimator and $BC3_p$ the bias corrected estimator proposed in this paper when the regressors are treated as predetermined.[22] The results show that uncorrected estimates of index coefficients are about 15 percent larger (in absolute value) 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).[23] The adjusted linear probability model gives estimates of the marginal effects closer to logit and probit estimates than the standard linear model estimates. 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 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 11 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 (in absolute value, 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.[24] The adjusted linear probability model, again, gives estimates of the average marginal effects closer to those obtained from logit and probit models than from 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%.

[22]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. How to extend the large-T bias corrections to the presence of time dummies is an open question in the large-T panel literature.

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

[24]Modified Maximum Likelihood estimates are taken from Carro (2007).

This elasticity is not sensitive to the omission of dynamics or to the bias corrections.

6 Summary and conclusions

This paper derives the analytical expression for the leading terms of large-T expansions of the biases of MLEs of index coefficients and average marginal effects in the fixed effects panel probit model. These expressions are used to construct bias corrections with improved finite-sample properties, to bound the bias for the estimates of the index coefficients, and to derive a small bias property for uncorrected fixed effects estimates of average marginal effects. It would be useful to know if this small bias property extends to other characteristics of the distribution of effects in the population, such as median effects or other quantile effects. Such analysis is expected to be more complicated because these characteristics are non-smooth functions of the data and therefore the standard expansions cannot be directly used. I leave this analysis for future research.

A Proofs of Main Results

A.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,*

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

and

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

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

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

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,

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

The second statement follows from differentiating both sides of expression (A.1) with respect to a. ■

A.2 Proof of Proposition 1

Proof. First, note that for the probit $f_{it} = \phi(X_{it}'\theta_0 + \alpha_i) := \phi_{it}$ and $g_{it} = -(X_{it}'\theta_0 + \alpha_i)\phi_{it}$. Then, substituting these expressions into the bias formulas (2.4)-(2.7) and dropping the terms that involve $\tilde{E}_T$ yields

$$\begin{aligned}
\beta_i &= \left\{E_T\left[H_{it}\phi_{it}X_{it}'\right]\theta_0\sigma_i^4 + \alpha_i\sigma_i^2\right\}/2,\\
\mathcal{J}_i &= E_T\left[H_{it}\phi_{it}X_{it}X_{it}'\right] - E_T\left[H_{it}\phi_{it}X_{it}\right]E_T\left[H_{it}\phi_{it}X_{it}'\right]\sigma_i^2, \quad \text{and}\\
b_i &= -\left\{E_T\left[H_{it}\phi_{it}X_{it}\right]E_T\left[H_{it}\phi_{it}X_{it}'\right]\theta_0\sigma_i^4 + E_T\left[H_{it}\phi_{it}X_{it}\right]\alpha_i\sigma_i^2\right\}/2\\
&+ \left\{E_T\left[H_{it}\phi_{it}X_{it}X_{it}'\right]\theta_0\sigma_i^2 + E_T\left[H_{it}\phi_{it}X_{it}\right]\alpha_i\sigma_i^2\right\}/2\\
&= \left\{\sigma_i^2\left(E_T\left[H_{it}\phi_{it}X_{it}X_{it}'\right] - E_T\left[H_{it}\phi_{it}X_{it}\right]E_T\left[H_{it}\phi_{it}X_{it}'\right]\sigma_i^2\right)\right\}\theta_0/2 = \sigma_i^2\mathcal{J}_i\theta_0/2.
\end{aligned}$$

Finally, we have for the bias

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

The second result follows since by condition (vii) $\sigma_i^2 = \sigma^2$ when $\alpha_i = \alpha \ \forall i$. ■

A.3 Proof of Proposition 2

Proof. We want to find a stochastic expansion for $\widehat{\mu} = \sum_{i=1}^n \widehat{\theta}\ \widehat{E}_T\left[\phi\left(X_{it}'\widehat{\theta} + \widehat{\alpha}_i(\widehat{\theta})\right)\right]/n$ as $n, T \to \infty$ and compare it to the population parameter of interest $\mu_0 = E_n\left[\theta_0 E_T\left[\phi\left(X_{it}'\theta_0 + \alpha_i\right)\right]\right]$.

First, note that by the Law of Large Number and Continuous Mapping Theorem, we have

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

as $n, T \to \infty$. 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:

$$\widehat{\xi}_{it}(\theta_T) = X_{it}'\theta_0 + \widehat{\alpha}_i(\theta_0) + \left[X_{it}' + \frac{\partial\widehat{\alpha}_i(\bar{\theta})}{\partial\theta'}\right](\theta_T - \theta_0), \tag{A.2}$$

where $\bar{\theta}$ lies between θ_T and θ_0. Using conditional independence across t given α_i and X_i^T, standard higher-order asymptotics for $\widehat{\alpha}_i(\theta_0)$ give (see, e.g., Ferguson, 1992, or Rilstone *et al.*, 1996), as $T \to \infty$

$$\widehat{\alpha}_i(\theta_0) = \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + R_{1i}, \quad \psi_i|X_i^T, \alpha_i \xrightarrow{d} \mathcal{N}(0, \sigma_i^2), \tag{A.3}$$

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 (taken from HK04). From the first order condition for $\widehat{\alpha}_i(\theta)$ and $\bar{\theta} \to \theta_0$ as $T \to \infty$, we have as $n, T \to \infty$

$$\frac{\partial\widehat{\alpha}_i(\bar{\theta})}{\partial\theta'} = -\frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]} + R_{2i} = \sigma_i^2 E_T\left[v_{it\theta}\right] + R_{2i}, \tag{A.4}$$

where $R_{2i} = O_p(T^{-1/2})$ and $E_T\left[R_{2i}\right] = O(T^{-1})$ uniformly in i, by the conditions of the proposition. Plugging (A.3) and (A.4) into the expansion for the index in (A.2) yields, for $\xi_{it} := X_{it}'\theta_0 + \alpha_i$,

$$\widehat{\xi}_{it}(\theta_T) = \xi_{it} + \psi_i/\sqrt{T} + \beta_{\xi_i}/T + R_{3i}, \tag{A.5}$$

where $\beta_{\xi_{it}} = \beta_i + T\left(X'_{it} + \sigma_i^2 E_T\left[v_{it\theta}\right]\right)(\theta_T - \theta_0)$, $R_{3i} = O_p(T^{-3/2})$, 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 of the static probit model (see the proof of Proposition 1) and $E_T\left[v_{it\theta}\right] = -E_T\left[H_{it}\phi_{it}X'_{it}\right]$, we have

$$\begin{aligned}
\beta_{\xi_{it}} &= \left(\sigma_i^4 E_T\left[H_{it}\phi_{it}X'_{it}\right]\theta_0 + \sigma_i^2\alpha_i\right)/2 + \left(X'_{it} - \sigma_i^2 E_T\left[H_{it}\phi_{it}X'_{it}\right]\right)\mathcal{B} + O(T^{-2}) \\
&= \sigma_i^2\xi_{it}/2 - \left(X'_{it} - \sigma_i^2 E_T\left[H_{it}\phi_{it}X'_{it}\right]\right)\sigma_i^2\theta_0/2 + \left(X'_{it} - \sigma_i^2 E_T\left[H_{it}\phi_{it}X'_{it}\right]\right)\mathcal{B} + O(T^{-2}) \\
&= \sigma_i^2\xi_{it}/2 - \mathcal{D}_{it} + O(T^{-2}),
\end{aligned}$$

where $\mathcal{D}_{it} = \left(X'_{it} - \sigma_i^2 E_T\left[H_{it}\phi_{it}X'_{it}\right]\right)(\mathcal{B}_i - \mathcal{B})$ and the remainder term is uniformly bounded in i. Substituting the expression for $\beta_{\xi_{it}}$ into (A.5) gives

$$\widehat{\xi}_{it}(\theta_T) = \left[1 + \sigma_i^2/2T\right]\xi_{it} + \psi_i/\sqrt{T} - \mathcal{D}_{it}/T + R_i,$$

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 $n, T \to \infty$,

$$\begin{aligned}
\widehat{\mu} &\xrightarrow{p} E_n E_T\left\{\theta_T E\left[\phi\left(\widehat{\xi}_{it}(\theta_T)\right)|X_i^T, \alpha_i\right]\right\} \\
&= E_n E_T\left\{\theta_T E\left[\phi\left(\left[1 + \sigma_i^2/2T\right]\xi_{it} + \psi_i/\sqrt{T} - \mathcal{D}_{it}/T\right)|X_i^T, \alpha_i\right]\right\} + O(T^{-2}) \\
&= E_n E_T\left\{\left(1 + \sigma_i^2/T\right)^{-1/2}\theta_T\phi\left(\frac{\left[1 + \sigma_i^2/2T\right]\xi_{it} - \mathcal{D}_{it}/T}{\sqrt{1 + \sigma_i^2/T}}\right)\right\} + O(T^{-2}) \\
&= \mu_0 + \frac{1}{T}E_n\left\{E_T\left[\phi(\xi_{it})\left(\xi_{it}\theta_0\left(X_{it} - \sigma_i^2 E_T\left[H_{it}\phi_{it}X_{it}\right]\right)' - \mathcal{I}_p\right)\right](\mathcal{B}_i - \mathcal{B})\right\} + O(T^{-2}) \\
&= \mu_0 + \frac{1}{T}\mathcal{B}_\mu + O\left(T^{-2}\right),
\end{aligned}$$

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

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

■

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 J. Hahn (2006), "A Likelihood-based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects," mimeo, CEMFI.

[9] Arellano, M., and J. Hahn (2007), "Understading Bias in Nonlinear Panel Models: Some Recent Developments," in R. Blundell, W. K. Newey and T. Persson, eds., *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Vol. 3*, Cambridge University Press: Cambridge.

[10] Carneiro, P., Hansen, K., and J. Heckman (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. (2007), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," " *Journal of Econometrics* 140(2), pp 503-528.

[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] 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.

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

[18] Dhaene, G., Jochmans, K., and B. Thuysbaert (2006), "Jackknife Bias Reduction for Nonlinear Dynamic Panel Data Models with Fixed Effects," mimeo, K.U. Leuven.

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

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

[21] Greene, W.H. (2004), “The Behavior of the Fixed Effects Estimator in Nonlinear Models,” *Econometrics Journal* 7, 98-119.

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

[23] 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.

[24] Hahn, J., and G. Kuersteiner (2004), “Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects,” unpublished manuscript. UCLA.

[25] Hahn, J., and G. Kuersteiner (2007), “Bandwidth Choice for Bias Estimators in Dynamic Panel Data Models,” mimeo, UCLA.

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

[27] Heckman, J. J. (1981), “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.

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

[29] 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.

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

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

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

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

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

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

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

[37] Mcfadden, D. (1974), “Conditional Logit Analysis of Qualitative Choice Analysis,” in *Frontiers in Econometrics,* ed. Zarembka. New York: Academic Press, 105-142.

[38] Mcfadden, D., and F. Reid (1975), “Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models,” *Transporation Research Record: Travel Behavior and Values* 534, 24-37.

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

[40] 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.

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

[42] 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.

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

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

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

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

[47] Rasch, G. (1970), *Probabilistic Models for Some Interlligence and Attainment Tests. Denmark Paedogiska,* Copenhagen.

[48] 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.

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

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

[51] 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.

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

[53] 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.

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

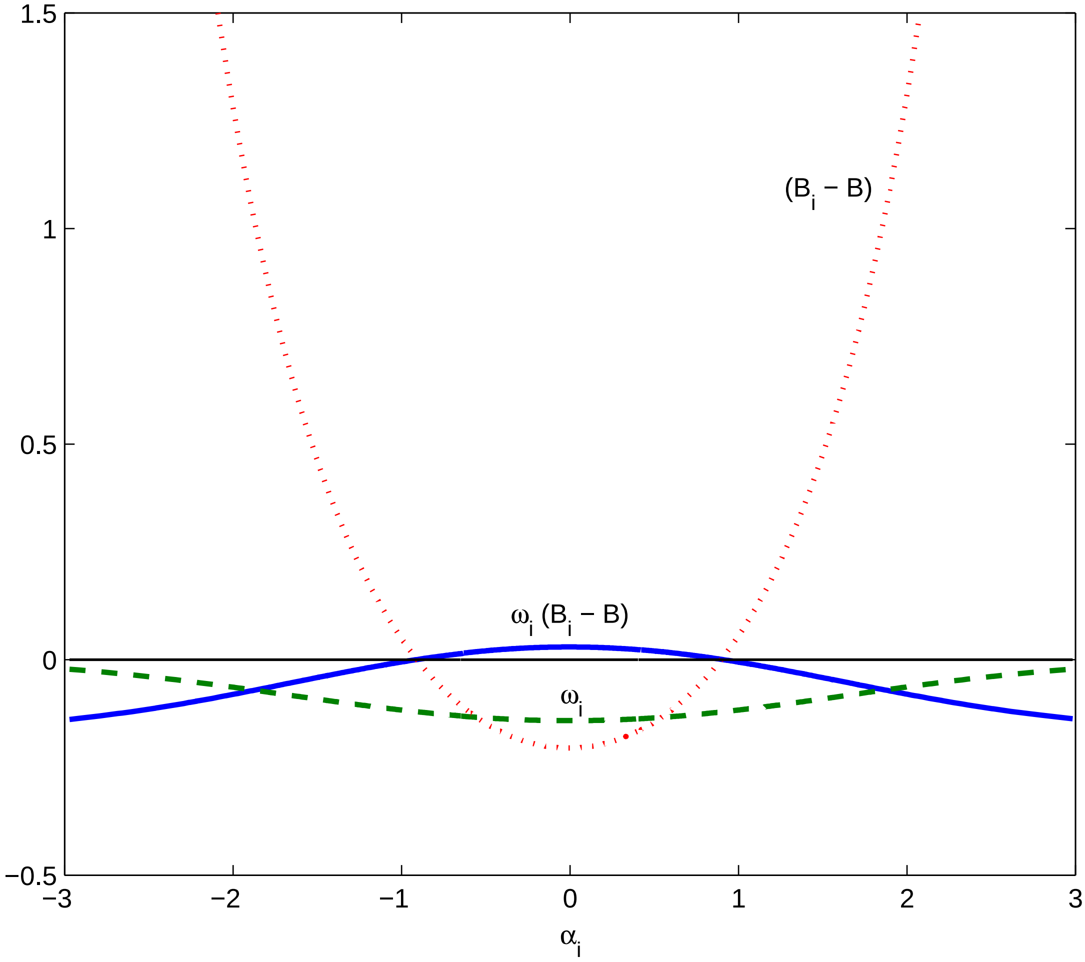

Figure 1: Components of the bias of the fixed effects estimator of the marginal effect: $\mathcal{B}_\mu = E_n[\omega_i(\mathcal{B}_i - \mathcal{B})]$. Individual effects and regressors are generated from independent standard normal distributions. Expressions are evaluated numerically using 1,000,000 replications.

Table 1: Asymptotic Bias (in parenthesis) and First Order Approximation for θ
(in percent of the true parameter value, $\theta_0 = 1$)

	Regressor				
Individual Effects	Nerlove	Normal	$\chi^2(1)$	$\chi^2(2)$	Bi(10,.9)
		T = 4			
Normal	27	27	27	27	28
	(39)	(40)	(40)	(38)	(41)
$\chi^2(1)$	24	24	22	22	26
	(36)	(35)	(35)	(36)	(36)
$\chi^2(2)$	24	25	23	23	26
	(37)	(36)	(36)	(36)	(37)
Bi(10,.9)	28	28	29	29	29
	(38)	(41)	(40)	(40)	(41)
C = 0	22	22	21	21	22
	(33)	(34)	(34)	(34)	(34)
		T = 8			
Normal	15	15	14	14	16
	(16)	(19)	(17)	(18)	(19)
$\chi^2(1)$	13	13	12	12	14
	(14)	(16)	(15)	(15)	(16)
$\chi^2(2)$	13	13	12	12	14
	(15)	(16)	(14)	(15)	(17)
Bi(10,.9)	15	15	15	15	17
	(17)	(20)	(18)	(18)	(21)
C = 0	13	13	12	12	13
	(13)	(17)	(15)	(16)	(17)

Notes: Bias formulae for first order approximation are evaluated numerically using 100,000 replications. The asymptotic bias is approximated by one simulation with n = 500,000 using Greene (2004) algorithm.

Table 2: Asymptotic Bias (in parenthesis) and First Order Approximation for μ
(in percent of the true parameter value, $\theta_0 = 1$)

	Regressor				
Individual Effects	Nerlove	Normal	$\chi^2(1)$	$\chi^2(2)$	Bi(10,.9)
		T = 4			
Normal	0	1	0	0	1
	(0)	(-1)	(0)	(-1)	(0)
$\chi^2(1)$	0	1	-1	-1	2
	(0)	(0)	(1)	(0)	(1)
$\chi^2(2)$	0	1	-1	-1	2
	(0)	(0)	(0)	(0)	(0)
Bi(10,.9)	1	1	2	1	1
	(-1)	(-1)	(0)	(-1)	(-1)
C = 0	0	0	0	0	0
	(1)	(0)	(1)	(0)	(0)
		T = 8			
Normal	0	0	0	-1	1
	(0)	(0)	(0)	(0)	(0)
$\chi^2(1)$	0	0	-1	-1	1
	(0)	(0)	(0)	(0)	(0)
$\chi^2(2)$	0	0	-1	-1	1
	(0)	(0)	(-1)	(0)	(0)
Bi(10,.9)	0	0	0	0	1
	(0)	-(1)	(0)	(-1)	(0)
C = 0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)

Notes: Bias formulae for first order approximation are evaluated numerically using 100,000 replications. The asymptotic bias is approximated by one simulation with n = 500,000 using Greene (2004) algorithm.

Table 3: 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's (2004) Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey's (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey's (2004) bias-corrected estimator based on general estimating equations; BC3 denotes the bias-corrected estimator proposed in this paper.

Table 4: 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's (2004) Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey's (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey's (2004) bias-corrected estimator based on general estimating equations; BC3 denotes the bias-corrected estimator proposed in this paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model.

Table 5: 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's (2004) bias-corrected estimator; HK denotes Honoré and Kyriazidou's (2000) bias-corrected estimator; MML denotes Carro's (2007) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in this paper. The Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 6: 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's (2004) bias-corrected estimator; HK denotes Honoré and Kyriazidou's (2000) bias-corrected estimator; MML denotes Carro's (2007) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in this paper. The Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 7: 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's (2004) bias-corrected estimator; HK denotes Honoré and Kyriazidou's (2000) bias-corrected estimator; MML denotes Carro's (2007) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in this paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model; BC-LPM denotes Nickell's (1981) bias-corrected adjusted linear probability model; BC-LPM-FS denotes Nickell's (1981) bias-corrected linear probability model. The Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 8: 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's (2004) bias-corrected estimator; HK denotes Honoré and Kyriazidou's (2000) bias-corrected estimator; MML denotes Carro's (2007) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in this paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model; BC-LPM denotes Nickell's (1981) bias-corrected adjusted linear probability model; BC-LPM-FS denotes Nickell's (1981) bias-corrected linear probability model. The Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 9: 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 10: Female Labor Force Participation (n = 1461, T = 9), Static Model

Estimator	PROBIT				LOGIT					LPM			
	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	-9.20	-9.73	-9.46	-11.19	-9.92	-11.72
	(0.70)	(0.71)	(0.70)	(0.77)	(0.72)	(0.72)	(0.72)	(0.72)	(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	-5.45	-5.72	-5.54	-6.09	-5.80	-6.38
	(0.66)	(0.66)	(0.66)	(0.71)	(0.67)	(0.67)	(0.67)	(0.67)	(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	-1.76	-1.94	-1.78	-1.25	-1.97	-1.45
	(0.53)	(0.53)	(0.53)	(0.59)	(0.54)	(0.54)	(0.54)	(0.54)	(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	-3.22	-3.17	-3.17	-3.61	-3.09	-3.59
	(0.70)	(0.69)	(0.70)	(0.74)	(0.71)	(0.71)	(0.71)	(0.71)	(0.75)	(0.72)	(0.71)	(0.74)	(0.74)

Notes: All specifications include time dummies and a quadratic function of age. FE denotes uncorrected fixed effects estimator; JK denotes Hahn and Newey's (2004) Jackknife bias-corrected estimator; BC3 denotes the bias-corrected estimator proposed in this paper; BC3p denotes the bias-corrected estimator proposed in this 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 11: 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					
$Participation_{t-1}$	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 (%)					
$Participation_{t-1}$	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 specifications include time dummies and a quadratic function of age. FE denotes uncorrected fixed effects estimator; BC3 denotes the bias-corrected estimator proposed in this paper; MML denotes Carro's (2007) Modified Maximum Likelihood estimator; LPM - FE denotes adjusted linear probability model (see text); LPM-BC denotes Nickell's (1981) bias-corrected adjusted linear probability model; LPM-FE-FS denotes linear probability model; LPM-BC-FS denotes Nickell's (1981) bias-corrected linear probability model. Column [3] is taken from Carro (2007). The specification in Carro (2007) also includes a lag of Kids 0-2 with an 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 is used as an initial condition.
Source: PSID 1979-1988.