Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects

Iván Fernández-Val
Department of Economics
MIT
JOB MARKET PAPER

January 14, 2005

Abstract

Fixed effects estimates of structural parameters in nonlinear panel models can be severely biased due to the incidental parameters problem. In this paper I show that the first term in a large-T expansion of the incidental parameters bias for probit fixed effects estimators of index coefficients is proportional to the true parameter value for general distributions of regressors and individual effects. This result allows me to derive a lower bound for the bias that depends only on the number of time periods of the panel. Proportionality is also used to show that the biases of ratios of coefficients and average marginal effects are identically zero in the absence of heterogeneity. Moreover, for a wide range of distributions of regressors and individual effects, numerical examples show that these biases are also very small. These results help explain previous Monte Carlo evidence for probit fixed effects estimates of index coefficients and marginal effects. Additional Monte Carlo examples suggest that the small bias property for fixed effects estimators of marginal effects holds for logit and linear probability models, and for the effects of exogenous variables in dynamic discrete choice models. The properties of logit and probit fixed effects estimates of model parameters and marginal effects are illustrated through an analysis of female labor force participation using data from the PSID. The results suggest that the significant biases in fixed effects estimates of model parameters do not contaminate the estimates of marginal effects in static models.

JEL Classification: C23; C25; J22.

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

Acknowledgment. I wish to thank Josh Angrist, Victor Chernozhukov, and Whitney Newey for encouragement and advice. I thank Jesús Carro for sharing his data. For suggestions and comments, I am grateful to Paula Bustos, Sylvain Chassang, Michael Jansson, and the participants to the MIT Econometrics lunch. All remaining errors are mine. I gratefully acknowledge financial support from Fundación Caja Madrid and Fundación Ramón Areces. Please send comments or suggestions to ifern12@mit.edu.

1 Introduction

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

Incidental parameters bias has been a longstanding problem in econometrics, but general bias correction methods have been developed only recently. Efforts in this direction include Lancaster (2000), Hahn and Kuersteiner (2001), Woutersen (2002), Arellano (2002), Alvarez and Arellano (2003), Carro (2003), Hahn and Kuersteiner (2003), and Hahn and Newey (2004). I refer to the approaches taken in these papers as providing 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.[1] The idea behind these methods is to expand the incidental parameters bias of the estimator in orders of magnitude of T, and to remove an estimate of the leading term of the bias from the estimator.[2] As a result, the adjusted estimator has a bias of order T^{-2}, whereas the bias of the initial estimator is of order T^{-1}. This approach aims to approximate the properties of estimators in applications that use panels of moderate length, such as the PSID or the Penn World Table, where the most important part of the bias is captured by the first term of the expansion.

The first contribution of this paper is to provide new correction methods for parametric binary choice models that attain the semiparametric efficiency bound of the bias estimation problem. The improvement comes from using the parametric structure of the model more inten-

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

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

sively than in previous studies by taking conditional moments of the bias, given the regressors and individual effects. The correction is then constructed based on the new formulae. This approach is similar to the use of the conditional information matrix in the estimation of asymptotic variances in maximum likelihood, instead of other alternatives, such as the sample average of the outer product of the scores or the sample average of the negative Hessian (Porter, 2002).[3] The adjustment presented here not only simplifies the correction by removing terms with zero conditional expectation, but also reduces the incidental parameter bias more effectively than other large-T corrections.

The second contribution of the paper is to derive a lower bound and a proportionality result for the bias of probit fixed effects estimators of model parameters. The lower bound depends uniquely upon the number of time periods of the panel, and is valid for general distributions of regressors and individual effects. According to this bound, for instance, the incidental parameters bias is at least 20 % for 4-period panels and 10 % for 8-period panels. Proportionality, on the other hand, establishes that probit fixed effect estimators of model parameters are biased away from zero when the regressor is scalar, providing a theoretical explanation for the numerical evidence in previous studies (see, for e.g., Greene, 2002). It also implies that fixed effects estimators of ratios of coefficients do not suffer from the incidental parameters bias in probit models in the absence of heterogeneity. These ratios are often structural parameters of interest because they can be interpreted as marginal rates of substitution in many economic applications.

Finally, the bias of fixed effects estimators of marginal effects in probit models is explored.[4] The motivation for this analysis comes from the 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)*?"[5] Wooldridge conjectures that the estimators of the marginal effects have reasonable properties. Here, using the expansion of the bias for the fixed effects estimators of model parameters, I characterize the analytical expression for the bias of these average marginal effects. As Wooldridge anticipated, this bias is negligible relative to the true average effect for a wide range of distributions of regressors and individual effects, and is identically zero in the absence of heterogeneity. This helps explain the small

[3]Porter (2002) also shows that the conditional information matrix estimator attains the semiparametric efficiency bound for the variance estimation problem.

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

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

biases in the marginal effects estimates that Hahn and Newey (2004) (HN henceforth) find in their Monte Carlo example.

The results presented in this paper are also consistent with Angrist's (2001) argument for cross-sectional limited dependent variable (LDV) models. Angrist argues that much of the difficulty with LDV models comes from a focus on structural parameters, such as latent index coefficients in probit models, instead of directly interpretable causal effects, such as average treatment effects (see also Wooldridge, 2002; Wooldridge, 2003; and Hahn, 2001). He recommends the use of simple linear models, where the structural parameters are directly linked to the effects of interest, just as if the outcomes were continuous. Here, I show that the same approach of focusing directly on causal effects rather than structural parameters also pays off in panel data models. However, unlike Angrist (2001), I use nonlinear models that incorporate the restrictions on the data support explicitly. These models are better suited for LDVs in cases where some regressors are continuous or the model is not fully saturated.

Monte Carlo examples show that adjusted logit and probit estimators of model parameters based on the new bias formulae have improved finite sample properties. In particular, these corrections remove more effectively the incidental parameters bias and provide estimators with smaller dispersion than previous methods. Accurate finite sample inference for model parameters and marginal effects is obtained from distributions derived under asymptotic sequences where $T/n^{1/3} \to \infty$ in static panels with 4 periods and dynamic panels with 8 periods. Results are also consistent with the small bias property of marginal effects for static models; they suggest that the property holds for the effects of exogenous variables in dynamic models, but not for the effects of lagged dependent variables. Simple linear probability models, in the spirit of Angrist (2001), also perform well in estimating average marginal effects.

The properties of probit and logit fixed effects estimators of model parameters and marginal effects are illustrated with an analysis of female labor force participation using 10 waves from the Panel Survey of Income Dynamics (PSID). The analysis here is motivated by similar studies in labor economics, where panel binary choice processes have been widely used to model female labor force participation decisions (see, e.g., Hyslop, 1999; Chay and Hyslop, 2000; and Carro, 2003). In particular, I show that fixed effects estimators, while biased for index coefficients, give very similar estimates to their bias corrected counterparts for marginal effects in static models. On the other hand, uncorrected fixed effects estimators are biased away from zero for both index coefficients and marginal effects of the fertility variables in dynamic models that account for true state dependence. In this case, the bias corrections presented here are effective reducing the incidental parameters problem.

The paper is organized as follows. Section 2 describes the panel binary choice model and its maximum likelihood estimator. Section 3 reviews existing solutions to the incidental parameters problem and proposes improved correction methods for binary choice models. Section 4 derives the proportionality result of the bias in static probit models. Section 5 analyzes the properties of probit fixed effects estimators of marginal effects. Section 6 extends the previous results to dynamic models. Monte Carlo results and the empirical application are given in Sections 7 and 8, respectively. Section 9 concludes with a summary of the main results.

2 The Model and Estimators

2.1 The Model

Given a binary response Y and a $p \times 1$ regressor vector X, consider the following data generating process

$$Y = \mathbf{1}\left\{X'\theta_0 + \alpha - \epsilon \geq 0\right\}, \tag{2.1}$$

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

2.2 Fixed Effects MLE

In economic applications, regressors and individual heterogeneity are correlated because regressors are decision variables and individual heterogeneity usually represents variation in tastes or technology. To avoid imposing any structure on this relationship, I adopt a fixed-effects approach and treat the sample realization of the individual effects $\{\alpha_i\}_{i=1,...,n}$ as parameters to be estimated, see Mundlak (1978), Lancaster (2000), Arellano and Honoré (2000), and Arellano (2003) for a similar interpretation of fixed effects estimators.[6]

[6]Note that Kiefer and Wolfowitz's (1956) consistency result does not apply to here, since no assumption is imposed on the distribution of the individual effects conditional on regressors.

To estimate the model parameters, a sample of the observable variables for individuals followed in subsequent periods of time $\{y_{it}, x_{it}\}_{t=1,...,T;\ i=1,..,n}$ is available, where i and t usually index individuals and time periods, respectively.[7] Then, assuming that ϵ follows a known distribution conditional on regressors and individual effects, typically normal or logistic, a natural way of estimating this model is by maximum likelihood.[8] Thus, if ϵ_{it} are i.i.d. conditional on $\bar{x}_i$ and α_i, with cdf $F_\epsilon(\cdot|\bar{X}, \alpha)$, the conditional log-likelihood for observation i at time t is[9]

$$l_{it}(\theta, \alpha_i) \equiv y_{it} \log F_{it}(\theta, \alpha_i) + (1 - y_{it}) \log(1 - F_{it}(\theta, \alpha_i)), \tag{2.2}$$

where $F_{it}(\theta, \alpha_i)$ denotes $F_\epsilon(x_{it}'\theta + \alpha_i|\bar{X} = \bar{x}_i, \alpha = \alpha_i)$, and the MLE of θ, concentrating out the α_i's, is the solution to

$$\widehat{\theta} \equiv \arg\max_\theta \sum_{i=1}^{n} \sum_{t=1}^{T} l_{it}(\theta, \widehat{\alpha}_i(\theta))/nT, \quad \widehat{\alpha}_i(\theta) \equiv \arg\max_\alpha \sum_{t=1}^{T} l_{it}(\theta, \alpha)/T. \tag{2.3}$$

2.3 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. The estimation error of these estimates introduces bias in the estimates of model parameters in nonlinear models. To see this, for $z_i \equiv (\bar{y}_i, \bar{x}_i)$ and any function $m(\bar{z}_i, \alpha_i)$, let $\bar{E}[m(\bar{z}_i, \alpha_i)] \equiv E_{\bar{X}\alpha}\left\{E_{\bar{Y}}\left[m(\bar{Z}, \alpha)|\bar{X}, \alpha\right]\right\}$, where the first expectation is taken with respect to the unknown joint distribution of $(\bar{X}, \alpha)$ and the second with respect to the known distribution of $\bar{Y}|\bar{X}, \alpha$.[10] Then, from the usual maximum likelihood properties, for $n \to \infty$ with T fixed,

$$\widehat{\theta} \xrightarrow{p} \theta_T, \ \ \theta_T \equiv \arg\max_\theta \bar{E}\left[\sum_{t=1}^{T} l_{it}(\theta, \widehat{\alpha}_i(\theta))/T\right]. \tag{2.4}$$

[7] In the sequel, for any random variable Z, z_{it} denotes observation at period t for individual i; $\bar{Z}$ denotes a random vector with T copies of Z; and $\bar{z}_i$ denotes an observation of $\bar{Z}$, i.e. $\{z_{i1}, ..., z_{iT}\}$.

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

[9] Following the common practice in fixed effects panel models, I will assume that the regressor vector $\bar{X}$ is strictly exogenous. See Arellano and Carrasco (2003) for an example of random effects estimator with predetermined regressors.

[10] When the observations are independent across time periods the conditional cdf of $\bar{Y}$ given $(\bar{X}, \alpha)$ can be factorized as $F(Y|\bar{X}, \alpha) \times ... \times F(Y|\bar{X}, \alpha)$. For dependent data, if $\bar{Y} = (Y_T, ..., Y_1)$, then the conditional cdf of $\bar{Y}$ given $(\bar{X}, Y_0, \alpha)$ factorizes as $F(Y_T|\bar{X}, Y_{T-1}, ..., Y_0, \alpha) \times ... \times F(Y_1|\bar{X}, Y_0, \alpha)$. The conditional cdf of Y can then be obtained from the conditional cdf of ϵ, which is assumed to be known.

When the true conditional log-likelihood of Y is $l_{it}(\theta_0, \alpha_i)$ generally $\theta_T \neq \theta_0$, but $\theta_T \rightarrow \theta_0$ as $T \rightarrow \infty$. For the smooth likelihoods considered here, $\theta_T = \theta_0 + \frac{\mathcal{B}}{T} + O\left(\frac{1}{T^2}\right)$ for some $\mathcal{B}$.[11] By asymptotic normality of the MLE, $\sqrt{nT}(\widehat{\theta} - \theta_T) \xrightarrow{d} N(0, -\mathcal{J}^{-1})$ as $n \rightarrow \infty$, and therefore

$$\sqrt{nT}(\widehat{\theta} - \theta_0) = \sqrt{nT}(\widehat{\theta} - \theta_T) + \sqrt{nT}\frac{\mathcal{B}}{T} + O\left(\sqrt{\frac{n}{T^3}}\right). \tag{2.5}$$

Here we can see that even if we let T grow at the same rate as n, that is $T = O(n)$, the limiting distribution of $\widehat{\theta}$ will not be centered at the true parameter value.

2.4 Large-T Approximation to the Bias

In moderate-length panels the most important part of the bias is captured by the first term of the expansion, $\mathcal{B}$. A natural way to reduce bias is therefore to remove a consistent estimate of this term from the MLE. To implement this procedure, however, we need an analytical expression for $\mathcal{B}$.[12] This expression can be characterized using a stochastic expansion of the fixed effects estimator in orders of T.

A little more notation is useful for describing this expansion. Let

$$u_{it}(\theta, \alpha) \equiv \frac{\partial}{\partial \theta} l_{it}(\theta, \alpha), \quad v_{it}(\theta, \alpha) \equiv \frac{\partial}{\partial \alpha} l_{it}(\theta, \alpha), \tag{2.6}$$

and additional subscripts denote partial derivatives, e.g. $u_{it\theta}(\theta, \alpha) \equiv \partial u_{it}(\theta, \alpha)/\partial \theta'$. Then, the first order condition for the concentrated problem can be expressed as

$$0 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it}(\widehat{\theta}, \widehat{\alpha}_i(\widehat{\theta})). \tag{2.7}$$

Expanding this expression around the true parameter value θ_0 yields

$$0 = \widehat{u}(\theta_0) + \widehat{\mathcal{J}}(\bar{\theta})(\widehat{\theta} - \theta_0) \quad \Rightarrow \quad \widehat{\theta} - \theta_0 = -\widehat{\mathcal{J}}(\bar{\theta})^{-1}\widehat{u}(\theta_0), \tag{2.8}$$

where $\bar{\theta}$ lies between $\widehat{\theta}$ and θ_0;

$$\widehat{u}(\theta_0) \equiv \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it}(\theta_0, \widehat{\alpha}_i(\theta_0)) \tag{2.9}$$

[11]To see this intuitively, note that MLEs are implicit smooth functions of sample means and use the following result (Lehmann and Casella, 1998). Let $X_1, ..., X_T$ be i.i.d. with $E[X_1] = \mu_X$, $Var[X_1] = \sigma_X^2$, and finite fourth moment; suppose h is a function of real variable whose first four derivatives exist for all $x \in I$, where I is an interval with $\Pr\{X_1 \in I\} = 1$; and $h^{iv}(x) \leq M$ for all $x \in I$, for some $M < \infty$. Then $E\left[h(\sum_{t=1}^{T} X_t/T)\right] = h(\mu_X) + \sigma_X^2 h''(\mu_X)/2T + O(T^{-2})$.

[12]Jackknife is an alternative bias correction method that does not require an analytical form for $\mathcal{B}$, see HN.

is the fixed effects estimating equation evaluated at the true parameter value, the expectation of which is generally different from zero because of the randomness of $\widehat{\alpha}(\theta_0)$ and determines the bias; and

$$\widehat{\mathcal{J}}(\theta) \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}(\theta,\widehat{\alpha}_i(\theta)) + u_{it\alpha}(\theta,\widehat{\alpha}_i(\theta))\frac{\partial\widehat{\alpha}_i(\theta)}{\partial\theta'}\right\} \tag{2.10}$$

is the Jacobian of $\widehat{u}(\theta)$.

For the estimators of the individual effects, the first order condition is $\sum_{t=1}^{T} v_{it}(\theta,\widehat{\alpha}_i(\theta))/T = 0$. Differentiating this expressions with respect to θ and $\widehat{\alpha}_i$ yields

$$\frac{\partial\widehat{\alpha}_i(\theta)}{\partial\theta'} = -\frac{\widehat{E}_T\left[v_{it\theta}(\theta,\widehat{\alpha}_i(\theta))\right]}{\widehat{E}_T\left[v_{it\alpha}(\theta,\widehat{\alpha}_i(\theta)\right]}, \tag{2.11}$$

where $\widehat{E}_T\left[f_{it}\right] \equiv \sum_{t=1}^{T} f_{it}/T$, for any function $f_{it} \equiv f(z_{it})$. Plugging this expression into (2.10) and taking (probability) limits as $n, T \to \infty$

$$\widehat{\mathcal{J}}(\bar{\theta}) \xrightarrow{p} \bar{E}\left[E_T\left[u_{it\theta}\right] - E_T\left[u_{it\alpha}\right]\frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]}\right] \equiv \mathcal{J}, \tag{2.12}$$

where $E_T\left[f_{it}\right] \equiv \lim_{T\to\infty}\sum_{t=1}^{T} f_{it}/T = E_Z\left[f_{it}|\alpha\right]$, by the Law of Large Numbers for i.i.d. sequences. For notational convenience the arguments are omitted when the expressions are evaluated at the true parameter value, i.e. $v_{it\theta} = v_{it\theta}(\theta_0,\alpha_i)$. Here, $-\mathcal{J}^{-1}$ gives the asymptotic variance of the fixed effect estimator of θ_0 under correct specification.

For the estimating equation, note that using independence across t, standard higher-order asymptotics for the estimator of the individual effects give (e.g., Ferguson, 1992, or Rilstone et al., 1996), as $T \to \infty$

$$\begin{aligned} \widehat{\alpha}_i &= \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + o_p(1/T),\ \ \psi_i = \sum_{t=1}^{T}\psi_{it}/\sqrt{T} \xrightarrow{d} \mathcal{N}(0,\sigma_i^2), && (2.13)\\ \psi_{it} &= \sigma_i^2 v_{it},\ \ \sigma_i^2 = -E_T\left[v_{it\alpha}\right]^{-1},\ \ \beta_i = \sigma_i^2\left\{E_T\left[v_{it\alpha}\psi_{it}\right] + \frac{1}{2}\sigma_i^2 E_T\left[v_{it\alpha\alpha}\right]\right\}. && (2.14) \end{aligned}$$

Then, expanding $\widehat{u}(\theta_0)$ around the α_i's, and assuming that orders in probability correspond to orders in expectation, we have as $n, T \to \infty$

$$\begin{aligned} T\widehat{u}(\theta_0) &= T\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{E}_T\left[u_{it}\right] + \widehat{E}_T\left[u_{it\alpha}(\widehat{\alpha}_i - \alpha)\right] + \widehat{E}_T\left[u_{it\alpha\alpha}(\widehat{\alpha}_i - \alpha)^2\right]/2 + o_p(1/T)\right\}\\ &\xrightarrow{p} \bar{E}\left\{0 + E_T\left[u_{it\alpha}\right]\beta_i + E_T\left[u_{it\alpha}\psi_{it}\right] + \frac{1}{2}\sigma_i^2 E_T\left[u_{it\alpha\alpha}\right]\right\} \equiv b. && (2.15) \end{aligned}$$

Finally, the (first term of the large-T expansion of the) asymptotic bias is

$$T(\widehat{\theta} - \theta_0) \xrightarrow{p} T(\theta_T - \theta_0) = -p\lim\widehat{\mathcal{J}}(\bar{\theta})^{-1}\ p\lim T\widehat{u}(\theta_0) = -\mathcal{J}^{-1}b \equiv \mathcal{B}. \tag{2.16}$$

3 Bias Corrections for Discrete Choice Panel Data Models

Large-T correction methods remove the bias of fixed effects estimators up to a certain order of magnitude in T. In particular, the incidental parameters bias is reduced from $O(T^{-1})$ to $O(T^{-2})$ as $T \to \infty$, and the asymptotic distribution is centered at the true parameter value if $T/n^{1/3} \to \infty$. 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

$$\sqrt{nT}(\widehat{\theta}^c - \theta_0) = \sqrt{nT}(\widehat{\theta}^c - \theta_T^c) + O\left(\sqrt{\frac{n}{T^3}}\right). \tag{3.1}$$

These methods can take different forms depending on whether the adjustment is made in the estimator, estimating equation, or objective function (log-likelihood). The first purpose of this section is to review the existing methods of bias correction, focusing on how these methods can be applied to panel binary choice models. The second purpose is to modify the corrections in order to improve their asymptotic and finite sample properties. Finally, I compare the different alternatives in a simple example.

3.1 Bias Correction of the Estimator

HN propose the following correction

$$\widehat{\theta}^1 \equiv \widehat{\theta} - \frac{\widehat{\mathcal{B}}}{T}, \tag{3.2}$$

where $\widehat{\mathcal{B}}$ is an estimator of $\mathcal{B}$. Since $\widehat{\mathcal{B}}$ generally depends on $\widehat{\theta}$, i.e., $\widehat{\mathcal{B}} = \widehat{\mathcal{B}}(\widehat{\theta})$, they also suggest to iterate the correction by solving $\widehat{\theta}^\infty = \widehat{\theta} - \widehat{\mathcal{B}}(\widehat{\theta}^\infty)$. To estimate $\mathcal{B}$, HN give two alternatives. The first alternative, only valid for the likelihood setting, is based on replacing derivatives for outer products in the bias formulae using Bartlett identities, and then estimate expectations using sample means. The second possibility replaces expectations for sample means using directly the bias formulae for general estimating equations derived in Section 2. These formulae rely only upon the unbiasedness of the estimating equation at the true value of the parameters and individual effects, and therefore are more robust to misspecification.

I argue here that the previous distinction is not very important for parametric discrete choice models, since the estimating equations are only valid under correct specification of the conditional distribution of ϵ. In other words, these estimating equations do not have a quasi-likelihood interpretation under misspecification. Following the same idea as in the estimation of asymptotic variances in MLE, I propose to take conditional expectations of the bias formulae (conditioning on the regressors and individual effects), and use the resulting expressions to construct the corrections.[13] These new corrections have optimality asymptotic properties. In particular, using Brown and Newey (1998) results for efficient estimation of expectations, it follows that the estimator of the bias proposed here attains the semiparametric efficiency bound for the bias estimation problem.[14] Intuitively, taking conditional expectations removes zero-mean terms of the bias formula that only add noise to the analog estimators.

To describe how to construct the correction from the new bias formulae, it is convenient to introduce some more notation. Let

$$\begin{aligned} F_{it}(\theta) &\equiv F_{\epsilon}(x_{it}'\theta + \widehat{\alpha}_i(\theta)|\bar{X}, \alpha), \quad f_{it}(\theta) \equiv f_{\epsilon}(x_{it}'\theta + \widehat{\alpha}_i(\theta)|\bar{X}, \alpha), && (3.3) \\ g_{it}(\theta) &\equiv f_{\epsilon}'(x_{it}'\theta + \widehat{\alpha}_i(\theta)|\bar{X}, \alpha), \quad H_{it}(\theta) \equiv \frac{f_{it}(\theta)}{F_{it}(\theta)\left(1 - F_{it}(\theta)\right)}, && (3.4) \end{aligned}$$

where f is the pdf associated with F, and f' is the derivative of f. Also, define

$$\widehat{\sigma}_i^2(\theta) \equiv \widehat{E}_T \left[H_{it}\left(\theta\right) f_{it}(\theta) \right]^{-1}, \quad \widehat{\psi}_{it}(\theta) \equiv \widehat{\sigma}_i^2(\theta) H_{it}\left(\theta\right) \left[y_{it} - F_{it}(\theta) \right]. \tag{3.5}$$

Here, $\widehat{\sigma}_i^2(\theta)$ and $\widehat{\psi}_{it}(\theta)$ are estimators of the asymptotic variance and influence function, respectively, obtained from a expansion of $\widehat{\alpha}_i(\theta)$ as T grows after taking conditional expectations, see (2.13) and the expressions in Appendix A. Let

$$\begin{aligned} \widehat{\beta}_i(\theta) &= -\widehat{\sigma}_i^4(\theta) \widehat{E}_T \left[H_{it}\left(\theta\right) g_{it}\left(\theta\right) \right] / 2, && (3.6) \\ \widehat{\mathcal{J}}(\theta) &= -\frac{1}{n} \sum_{i=1}^{n} \left\{ \widehat{E}_T \left[H_{it}(\theta) f_{it}(\theta) x_{it} x_{it}' \right] - \widehat{\sigma}_i^2(\theta) \widehat{E}_T \left[H_{it}(\theta) f_{it}(\theta) x_{it} \right] \widehat{E}_T \left[H_{it}\left(\theta\right) f_{it}(\theta) x_{it}' \right] \right\}, && (3.7) \end{aligned}$$

where $\widehat{\beta}_i(\theta)$ is an estimator of the higher-order asymptotic bias of $\widehat{\alpha}_i(\theta)$ from a stochastic expansion as T grows, and $\widehat{\mathcal{J}}\left(\theta\right)$ is an estimator of the Jacobian of the estimating equation for θ. Then, the estimator of $\mathcal{B}$ is

$$\widehat{\mathcal{B}}(\theta) = -\widehat{\mathcal{J}}(\theta)^{-1} \widehat{b}(\theta), \;\; \widehat{b}(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \left\{ \widehat{E}_T \left[H_{it}(\theta) f_{it}(\theta) x_{it} \right] \widehat{\beta}_i(\theta) + \widehat{E}_T \left[H_{it}(\theta) g_{it}(\theta) x_{it} \right] \widehat{\sigma}_i^2(\theta)/2 \right\}, \tag{3.8}$$

where $\widehat{b}(\theta)$ is an estimator of the bias of the estimating equation of θ.

One step bias corrected estimators can then be formed by evaluating the previous expression at the MLE, that is $\widehat{\mathcal{B}} = \widehat{\mathcal{B}}(\widehat{\theta})$, and the iterated bias corrected estimator is the solution to

[13] Appendix A gives the expressions of the bias for discrete choice models after taking conditional expectations.

[14] Brown and Newey (1998) results apply here by noting that the bias formula can be decomposed in unconditional expectation terms. The efficient estimators for each of these terms is the corresponding sample analog of the conditional expectation, given the regressors and individual effects. Then, the argument follows by delta method. See also Porter (2002).

$\widehat{\theta}^{\infty} = \widehat{\theta} - \widehat{\mathcal{B}}(\widehat{\theta}^{\infty})$. Monte Carlo experiments in Section 7 show that the previous adjustments improve the finite sample performance of the corrections.

3.2 Bias Correction of the Estimating Equation

The source of incidental parameters bias is the non-zero expectation of the estimating equation (first order condition) for $\widehat{\theta}$ at the true parameter value θ_0, see (2.15). This suggests an alternative correction consisting of a modified estimating equation that has no bias at θ_0, up to a certain order in T (see, for e.g., Woutersen, 2002; HN; and Fernández-Val, 2004).[15] For the discrete choice model, the score-corrected estimator is the solution to

$$0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}(\tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})) - \frac{1}{T}\widehat{b}(\tilde{\theta}). \tag{3.9}$$

HN and Fernández-Val (2004) show that this method is equivalent to the iterated bias correction of the estimator when the initial estimating equation is linear in θ. In general, the iterated estimator is the solution to an approximation to the unbiased estimating equation.

3.3 Modified (Profile) Maximum Likelihood (MML)

Cox and Reid (1987), in the context of robust inference with nuisance parameters, develop a method for reducing the sensitivity of MLEs of structural parameters to the presence of incidental parameters.[16] This method consists of adjusting the likelihood function to reduce the order of the bias of the corresponding estimating equation (see Liang, 1987; McCullagh and Tibsharani, 1990; and Ferguson, Reid and Cox, 1991). Lancaster (2000) and Arellano (2003) show that the modified profile likelihood, i.e. concentrating out the α_i's, takes the following form for panel discrete choice models

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} l_{it}(\theta, \widehat{\alpha}_i(\theta)) - \frac{1}{2T}\frac{1}{n}\sum_{i=1}^{n}\log\frac{-\widehat{E}_T\left[v_{it\alpha}(\theta, \widehat{\alpha}_i(\theta))\right]}{\widehat{E}_T\left[H_{it}(\theta) f_{it}(\theta)\right]^2} + \frac{1}{2}\log T. \tag{3.10}$$

Appendix D shows that the estimating equation of the modified profile likelihood is equivalent to (3.9), up to order $o_p(1/T)$. The difference with (3.9) is that the MML estimating equation does not use conditional expectations of all the terms. For the logit, however, the two methods exactly coincide since the correction factor of the likelihood does not depend on ϵ. In this case

[15] Neyman and Scott (1948) suggest this method, but do not give the general expression for the bias of the estimating equation.

[16] "Roughly speaking 'nuisance' parameters are those which are not of primary interest; 'incidental' parameters are nuisance parameters whose number increases with the sample size." C.f., Lancaster (2000), footnote 10.

$\widehat{E}_T\left[v_{it\alpha}(\theta,\widehat{\alpha}_i(\theta))\right]=-\widehat{E}_T\left[H_{it}(\theta)f_{it}(\theta)\right]$, $H_{it}(\theta)=1$, and the modified likelihood takes the simple form

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}l_{it}(\theta,\widehat{\alpha}_i(\theta))+\frac{1}{2T}\frac{1}{n}\sum_{i=1}^{n}\log\widehat{E}_T\left[f_{it}(\theta)\right]+\frac{1}{2}\log T. \tag{3.11}$$

3.4 Example: Andersen (1973) Two-Period Logit Model

The previous discussion suggests that the most important distinction between the correction methods is whether to adjust the estimator or the estimating equation, and in the former case whether to use a one-step or an iterative procedure. Here, I compare the asymptotic properties of these alternative procedures in a simple example from Andersen (1973). This example is convenient analytically because the fixed-T probability limit of the MLE has a closed-form expression. This expression allows me to derive the probability limits of the bias-corrected estimators, and to compare them to the true parameter value.

The model considered is

$$y_{it}=1\left\{x_{it}\theta_0+\alpha_i-\epsilon_{it}\geq 0\right\},\epsilon_{it}\sim\mathcal{L}(0,\pi^2/3)\quad t=1,2;\ i=1,...,n, \tag{3.12}$$

where $x_{i1}=0$ and $x_{i2}=1$ for all i, and $\mathcal{L}$ denotes the standardized logistic distribution. Andersen (1973) shows that the MLE, $\widehat{\theta}$, converges to $2\theta_0\equiv\theta^{ML}$, as $n\to\infty$, and derives a fixed-T consistent estimator for this model, the conditional logit estimator. Using the probability limit of the MLE and the expression of the bias in (3.8), the limit of the one-step bias-corrected estimator is

$$\widehat{\theta}^1=\widehat{\theta}-\left(e^{\widehat{\theta}/2}-e^{-\widehat{\theta}/2}\right)/2\xrightarrow{p}2\theta-\left(e^{\theta}-e^{-\theta}\right)/2\equiv\theta^1. \tag{3.13}$$

For the iterated bias-corrected estimator, the limit of the estimator (θ^{∞}) is the solution to the following nonlinear equation

$$\theta^{\infty}=2\theta-\left(e^{\theta^{\infty}/2}-e^{-\theta^{\infty}/2}\right)/2. \tag{3.14}$$

Finally, Arellano (2003) derives the limit for the score-corrected estimator[17]

$$\tilde{\theta}\xrightarrow{p}2\log\left(\frac{5e^{\theta}+1}{5+e^{\theta}}\right)\equiv\theta^s. \tag{3.15}$$

Figures 1 and 2 plot the limit of the corrected estimators as functions of the true parameter value θ_0.[18] Here, we can see that the large-T adjustments produce significant improvements over

[17] He obtains this result for the MML estimator, but MML is the same as score correction for the logit case.

[18] See Arellano (2002) for a similar exercise comparing the probability limits of MML and ML estimators.

the MLE for a wide range of parameter values, even for $T = 2$. Among the corrected estimators considered, no estimator uniformly dominates the rest in terms of having smaller bias for all the parameter values. Thus, the one-step bias correction out-performs the other alternatives for low values of θ, but its performance deteriorates very quickly as the true parameter increases; the score correction dominates for medium range parameter values; and the iterated correction becomes the best for high values of θ.

4 Bias for Static Panel Probit Model

The expression for the bias takes a simple form for the static panel probit model, which helps explain the results of previous Monte Carlo studies (Greene, 2002; and HN). In particular, the bias can be expressed as a matrix-weighted average of the true parameter value, where the weighting matrices are positive definite. This implies that probit fixed effects estimators are biased away from zero if the regressor is scalar (as in the studies aforementioned). This property also holds regardless of the dimension of the regressor vector in the absence of heterogeneity, because in this case the weighting matrix is a scalar multiple of the identity matrix (Nelson, 1995). In general, however, matrix-weighted averages are difficult to interpret and sign except in special cases (see Chamberlain and Leamer, 1976). These results are stated in the following proposition:

Proposition 1 (Bias for Model Parameters) *Assume that (i)* $\epsilon_{it}|\bar{X}_i, \alpha_i \sim$ *i.i.d.* $\mathcal{N}(0,1)$*, (ii)* $E[XX'|\alpha]$ *exists and is nonsingular for almost all* α*, (iii)* $E\left[\|X\|^{12}|\alpha\right] < \infty$ *for almost all* α*, and (iv)* $n = o(T^3)$.[19] *Then,*

1.

$$\mathcal{B} = \frac{1}{2}\bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\theta_0, \tag{4.1}$$

where

$$\begin{aligned}\mathcal{J}_i &= E_T\left[H_{it}f_{it}x_{it}x_{it}'\right] - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]E_T\left[H_{it}f_{it}x_{it}'\right], &(4.2)\\ \sigma_i^2 &= E_T\left[H_{it}f_{it}\right]^{-1} = E_T\left\{\phi(x_{it}'\theta_0+\alpha_i)^2/\left[\Phi(x_{it}'\theta_0+\alpha_i)\left(1-\Phi(x_{it}'\theta_0+\alpha_i)\right)\right]\right\}^{-1}. &(4.3)\end{aligned}$$

2. $\bar{E}\left[\mathcal{J}_i\right]^{-1}\sigma_i^2\mathcal{J}_i$ *is positive definite for almost all* α_i*.*

3. If $\alpha_i = \alpha\ \forall i$*, then*

$$\mathcal{B} = \frac{1}{2}\sigma^2\theta_0, \tag{4.4}$$

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

[19] $\|\cdot\|$ denotes the Euclidean norm.

Proof. See Appendix C. ■

Condition (i) is the probit modeling assumption; condition (ii) is standard for MLE (Newey and McFadden, 1994), and guarantees identification and asymptotic normality for MLEs of model parameters and individual effects; and assumptions (iii) and (iv) guarantee the existence of, and uniform convergence of remainder terms in, the higher-order expansion of the bias (HN, and Fernández-Val, 2004). Note that the second result follows because σ_i^2 is the asymptotic variance of the estimator of the individual effect α_i, and $\mathcal{J}_i$ corresponds to the contribution of individual i to the inverse of the asymptotic variance of the estimator of the model parameter θ_0. Moreover, since $\sigma_i^2 \geq \Phi(0)\left[1-\Phi(0)\right]/\phi(0)^2 = \pi/2$, $\mathcal{B}$ can be bounded from below.

Corollary 1 *Under the conditions of Proposition 1*

$$\|\mathcal{B}\| \geq \frac{\pi}{4}\ \|\theta_0\|. \tag{4.5}$$

When the regressor is scalar or there is no heterogeneity, this lower bound establishes that the first order bias for each index coefficient is at least $\pi/8 \approx 40\%$, $\pi/16 \approx 20\%$ and $\pi/32 \approx 10\%$ for panels with 2, 4 and 8 periods, respectively.[20] In general, these bounds apply to the norm of the coefficient vector. Tighter bounds can be also established for the proportionate bias, $\|\mathcal{B}\|/\|\theta_0\|$, as a function of the true parameter value θ_0. These bounds, however, depend on the joint distribution of regressor and individual effects, and are therefore application specific. Thus, using standard matrix algebra results (see, e.g., Rao, 1973, p. 74), the proportionate bias can be bounded from below and above by the minimum and maximum eigenvalues of the matrix $\bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]/2$, for any value of the parameter vector θ_0.[21]

The third result of the proposition establishes that the bias is proportional to the true parameter value in the absence of heterogeneity. Proportionality implies, in turn, zero bias for fixed effects estimators of ratios of index coefficients. These ratios are often structural parameters of interest because they are direct measures of the relative effect of the regressors, and can be interpreted as marginal rates of substitution in many economic applications.

Corollary 2 *Assume that the conditions of the Proposition 1 hold and $\alpha_i = \alpha \ \ \forall i$. Then, for*

[20] For two-period panels, the incidental parameters bias of the probit estimator is 100 % (Heckman, 1981). Part of the difference between the bias and the lower bound in this case can be explained by the importance of higher order terms, which have growing influence as the number of periods decreases.

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

any $j \neq k \in \{1, ..., p\}$ *and* $\theta = (\theta_1, ..., \theta_p)$

$$\frac{\widehat{\theta}_j}{\widehat{\theta}_k} \xrightarrow{p} \frac{\theta_{0,j}}{\theta_{0,k}} + O(T^{-2}). \tag{4.6}$$

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

5 Marginal Effects: Small Bias Property

5.1 Parameters of Interest

In discrete choice models the ultimate quantities of interest are often the marginal effects of specific changes in the regressors on the response conditional probability (see, e.g., Angrist, 2001; Ruud 2001; Greene, 2002; Wooldridge, 2002; and Wooldridge, 2003). However, unlike in linear models, structural parameters in nonlinear index models are only informative about the sign and relative magnitude of the effects. In addition, an attractive feature of these models is that marginal effects are heterogeneous across individuals. This allows, for instance, the marginal effects to be decreasing in the propensity (measured by the individual effect) to experience the event. Thus, individuals more prone to work are arguably less sensitive to marginal changes on other observable characteristics when deciding labor force participation.

For a model with two regressors, say X_1 and X_2, and corresponding parameters θ_1 and θ_2, the marginal effect of a one-unit increase in X_1 on the conditional probability of Y is defined as

$$F_\epsilon((x_1 + 1)\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha). \tag{5.1}$$

When X_1 is continuous, the previous expression is usually approximated by a local version based on the derivative of the conditional probability with respect to x_1, that is

$$\frac{\partial}{\partial x_1} F_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) = \theta_1 f_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha), \tag{5.2}$$

where $f_\epsilon(\cdot|\bar{X}, \alpha)$ is the conditional pdf associated with $F_\epsilon(\cdot|\bar{X}, \alpha)$. These measures are heterogenous in the individual effect α and the level chosen for evaluating the regressors.

What are the relevant effects to report? A common practice is to give some summary measure, for example, the average effect or the effect for some interesting value of the regressors.

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

$$\mu(x_1) = \int \left[F_\epsilon((x_1+1)\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) \right] dG_{\bar{X}_2,\alpha}(\bar{x}_2, a), \tag{5.3}$$

or

$$\mu = \int \theta_1 f_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) dH_{\bar{X}_1,\bar{X}_2,\alpha}(\bar{x}_1, \bar{x}_2, a), \tag{5.4}$$

where G and H are the joint distribution of $(\bar{X}_2, \alpha)$ and $(\bar{X}, \alpha)$, respectively, and x_1 is some interesting value of X_1. The previous measures correspond to different thought experiments. The first measure, commonly used for discrete variables, corresponds to the counterfactual experiment where the change on the outcome probability is evaluated as if all the individuals would have chosen x_1 initially and receive an additional unit of X_1. The second measure, usually employed for continuous variables, is the average derivative of the response probabilities with respect to x_1, i.e., the average effect of giving one additional unit of X_1. The fixed effects estimators for these measures are

$$\widehat{\mu}(x_1) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[F_\epsilon((x_1+1)\widehat{\theta}_1 + x_{2it}\widehat{\theta}_2 + \widehat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_1\widehat{\theta}_1 + x_{2it}\widehat{\theta}_2 + \widehat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha) \right], \tag{5.5}$$

and

$$\widehat{\mu} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \widehat{\theta}_1 f_\epsilon(x_{1it}\widehat{\theta}_1 + x_{2it}\widehat{\theta}_2 + \widehat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha), \tag{5.6}$$

respectively. Note that the first measure corresponds to the average treatment effect if X_1 is a treatment indicator.

These effects can be calculated also for subpopulations of interest by conditioning on the relevant values of the covariates. For example, if X_1 is binary (treatment indicator), the average treatment effect on the treated (ATT) is

$$\mu_{ATT} = \int \left[F_\epsilon(\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) \right] dG_{\bar{X}_2,\alpha}(x_2, a|X_1 = 1), \tag{5.7}$$

and can be estimated by

$$\widehat{\mu}_{ATT} = \frac{1}{N_1} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[F_\epsilon(\widehat{\theta}_1 + x_{2it}\widehat{\theta}_2 + \widehat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha) - F(x_{2it}\widehat{\theta}_2 + \widehat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha) \right] \mathbf{1}\left\{x_{1it} = 1\right\}, \tag{5.8}$$

where $N_1 = \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}\left\{x_{1it} = 1\right\}$. Other alternative measures used in cross-section models, such as the effect evaluated for an individual with average characteristics, are less attractive for panel data models because they raise conceptual and implementation problems (see Carro, 2003, for a related discussion about other measures of marginal effects).[22]

5.2 Bias Correction of Marginal Effects

HN develop analytical and jackknife bias correction methods for fixed effect averages, which include marginal effects. Let $m(r, \theta, \alpha)$ denote the change in the outcome conditional probability as a response to a one-unit increase in the first regressor $F_\epsilon\left((r_1+1)\theta_1 + r_{-1}'\theta_{-1} + \alpha | \bar{X}, \alpha\right) - F_\epsilon(r_1\theta_1 + r_{-1}'\theta_{-1} + \alpha | \bar{X}, \alpha)$, or its local approximation $\theta_1 f_\epsilon(r'\theta + \alpha | \bar{X}, \alpha)$ if X_1 is continuous.[23] The object of interest is then

$$\mu = \bar{E}\left[m(r, \theta_0, \alpha)\right], \tag{5.9}$$

where $r = (x_1, X_2)$ for discrete X_1 and $r = X$ for continuous X_1.[24] The fixed effects MLE of μ is then given by

$$\widehat{\mu} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m\left(r_{it}, \widehat{\theta}, \widehat{\alpha}_i(\widehat{\theta})\right), \tag{5.10}$$

where $r_{it} = (x_1, x_{2it})$ or $r_{it} = x_{it}$. For the bias corrections, let $\tilde{\theta}$ be a bias-corrected estimator (either one-step, iterated or derived from a bias-corrected estimating equation) of θ_0 and $\tilde{\alpha}_i = \widehat{\alpha}_i(\tilde{\theta})$, $(i = 1, ..., n)$, the corresponding estimators of the individual effects.[25] Then, a bias-corrected estimator of μ is given by

$$\tilde{\mu} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m\left(r_{it}, \tilde{\theta}, \tilde{\alpha}_i\right) - \frac{1}{T}\tilde{\Delta}, \tag{5.11}$$

$$\tilde{\Delta} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{ m_\alpha\left(r_{it}, \tilde{\theta}, \tilde{\alpha}_i\right)\left[\widehat{\beta}_i\left(\tilde{\theta}\right) + \widehat{\psi}_{it}\left(\tilde{\theta}\right)\right] + \frac{1}{2} m_{\alpha\alpha}\left(r_{it}, \tilde{\theta}, \tilde{\alpha}_i\right)\widehat{\sigma}_i^2\left(\tilde{\theta}\right)\right\}, \tag{5.12}$$

where subscripts on m denote partial derivatives. Note that the $\widehat{\psi}_{it}\left(\tilde{\theta}\right)$ term can be dropped since r_{it} does not depend on ϵ_{it}.

[22] On the conceptual side, Chamberlain (1984) and Ruud (2000) argue that this effect may not be relevant for most of the population. The practical obstacle relates to the difficulty of estimating average characteristics in panel models. Thus, replacing population expectations for sample analogs does not always work in binary choice models estimated using a fixed-effects approach. The problem here is that the MLEs of the individual effects are unbounded for individuals that do not change status in the sample, and therefore the sample average of the estimated individual effects is generally not well defined.

[23] For a $p \times 1$ vector v, v_i denotes the $i-th$ component and v_{-i} denotes $(v_1, ..., v_{i-1}, v_{i+1}, ..., v_p)'$.

[24] If X_1 is binary r_1 is usually set to 0.

[25] Bias-corrected estimators for marginal effects can also be constructed from fixed-T consistent estimators. Thus, the conditional logit estimator can be used as $\tilde{\theta}$ in the logit model. This possibility is explored in the Monte Carlo experiments and the empirical application.

5.3 Panel Probit: Small Bias Property

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

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

$$\widehat{\mu} \xrightarrow{p} \mu + \frac{1}{T}\mathcal{B}_{\mu} + O(T^{-2}), \tag{5.13}$$

where

$$\mathcal{B}_{\mu} = \frac{1}{2}\bar{E}\left\{\phi\left(\xi_{it}\right)\left[\xi_{it}\theta_0\left(x_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]\right)' - \mathcal{I}_p\right]\left(\sigma_i^2\mathcal{I}_p - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\right)\right\}\theta_0, \tag{5.14}$$

$$\xi_{it} = x_{it}'\theta_0 + \alpha_i, \tag{5.15}$$

$$\sigma_i^2 = E_T\left[H_{it}f_{it}\right]^{-1} = E_T\left\{\phi(\xi_{it})^2/\left[\Phi(\xi_{it})\Phi(-\xi_{it})\right]\right\}^{-1}, \tag{5.16}$$

$$\mathcal{J}_i = -\left\{E_T\left[H_{it}f_{it}x_{it}x_{it}'\right] - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]E_T\left[H_{it}f_{it}x_{it}'\right]\right\}, \tag{5.17}$$

and $\mathcal{I}_p$ *denotes a* $p \times p$ *identity matrix.*

Proof. See Appendix C. ■

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

$$\mathcal{B}_{\mu} = E_{\alpha}\left[\mathcal{B}_{\mu_i}\right] = E_{\alpha}\left[\delta_i\pi_i\right]\theta_0/2, \tag{5.18}$$

with $\delta_i = E_X\left\{\phi(\xi_{it})\left[\xi_{it}\left(\xi_{it} - E_T\left[H_{it}f_{it}\right]^{-1}E_T\left[H_{it}f_{it}\xi_{it}\right]\right) - 1\right]|\alpha\right\}$, $\pi_i = \sigma_i^2 - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]$, and $\mathcal{J}_i = -\left\{E_T\left[\omega_{it}\xi_{it}^2\right] - \sigma_i^2 E_T\left[\omega_{it}\xi_{it}\right]^2\right\}$. Here, we can see that the bias is an even function of the index, ξ_{it}, since all the terms are centered around weighted means that are symmetric around the origin.[26] The terms δ_i and π_i have U shapes and take the same sign for $\alpha_i = 0$ (when the mean of the regressors is absorbed in the individual effects); the term $\phi(\xi_{it})$ in δ_i acts to

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

reduce the weights in the tails, where the other components are large. This is shown in Figure 3, which plots the components of the bias for independent normally distributed regressor and individual effect. In this case the bias, as a function of α, is positive at zero and takes negative values as we move away from the origin. Then, positive and negative values compensate each other when they are integrated using the distribution of the individual effect to obtain $\mathcal{B}_\mu$.

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

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

$$\widehat{\mu} \xrightarrow{p} \mu + O_p(T^{-2}). \tag{5.19}$$

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

$$\widehat{\mu} \xrightarrow{p} \bar{E}\left\{\theta_T E_T\left[\phi\left(x_{it}'\theta_T + \widehat{\alpha}_i(\theta_T)\right)\right]\right\}. \tag{5.20}$$

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

$$\theta_T = \theta_0 + \frac{1}{2T}\sigma^2\theta_0 + O\left(T^{-2}\right), \ \widehat{\alpha}_i(\theta_T) = \alpha + \tilde{\psi}_i/\sqrt{T} + \alpha\sigma^2/2T + R_i/T^{3/2}, \tag{5.21}$$

where under the conditions of Proposition 2

$$\tilde{\psi}_i \overset{a}{\sim} N(0, \sigma^2), \ \sigma^2 = E_T\left[H_{it} f_{it}\right]^{-1}, \quad \bar{E}\left[R_i\right] = O(T^{-2}). \tag{5.22}$$

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

$$\widehat{\xi}_{it} = \left(1 + \sigma^2/2T\right)\xi_{it} + \psi_i/\sqrt{T} + R_i/\sqrt{T} + O(T^{-2}), \quad \xi_{it} = x_{it}'\theta_0 + \alpha. \tag{5.23}$$

Finally, replacing this expression in (5.20), using the convolution properties of the normal distribution, see Lemma 1 in Appendix C, and assuming that orders in probability correspond to orders in expectation,

$$\begin{aligned} \widehat{\mu} \quad &\xrightarrow{p} \quad \bar{E}\left\{\theta_T \ E\left[\phi(\widehat{\xi}_{it})|\bar{X}, \alpha\right]\right\} = \bar{E}\left[\theta_T \ \int \phi\left(\left(1 + \frac{\sigma^2}{2T}\right)\xi_{it} + v\right)\frac{\sqrt{T}}{\sigma} \ \phi\left(\frac{\sqrt{T}v}{\sigma}\right)dv\right] + O(T^{-2}) \\ &= \quad \bar{E}\left[\frac{\left(1 + \sigma^2/2T\right)}{\sqrt{1 + \sigma^2/T}} \ \theta_0 \ \phi\left(\frac{\left(1 + \sigma^2/2T\right)\xi}{\sqrt{1 + \sigma^2/T}}\right)\right] + O(T^{-2}) = \mu + O(T^{-2}), \end{aligned} \tag{5.24}$$

since by a standard Taylor expansion

$$\left(1 + \sigma^2/T\right)^{-1/2}\left(1 + \sigma^2/2T\right) \quad = \quad \left(1 - \frac{\sigma^2}{2T} + O(T^{-2})\right)\left(1 + \frac{\sigma^2}{2T}\right) = 1 + O(T^{-2}). \tag{5.25}$$

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

The intuition for this result is the equivalent for panel probit of the consistency of average survivor probabilities in the linear Gaussian panel model. Thus, HN show that $\widehat{S} = \sum_{i=1}^{n} \Phi\left(\frac{x'\widehat{\theta}+\widehat{\alpha}_i(\widehat{\theta})}{\widehat{\sigma}}\right)/n$ is a consistent estimator for $S = \bar{E}\left\{\Phi\left(\frac{x'\theta_0+\alpha_i}{\sigma}\right)\right\}$ for fixed T, because averaging across individuals exactly compensates the bias of the estimator of σ. In the nonlinear model, however, the result holds only approximately, since averaging reduces the bias of the MLE of the average effects by one order of magnitude from $O\left(T^{-1}\right)$ to $O\left(T^{-2}\right)$.

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

In general, the bias depends upon the degree of heterogeneity and the joint distribution of regressors and individual effects. Table 1 reports numerical values for the bias of fixed effects estimators of model parameters and marginal effects (in percent of the true value) for several distributions of regressors and individual effects. These examples correspond to an 8-period model with one regressor, and the model parameter θ_0 equal to 1. All the distributions, except for the Nerlove process for the regressor, are normalized to have zero mean and unit variance.[27] The numerical results show that the first term of the bias for the marginal effect is below 2% for all the configurations considered, and is always lower than the bias for the model parameter, which is about 15% (larger than the lower bound of 10%). The values of the bias for Nerlove regressor and normal individual effect are close to their Monte Carlo estimates in Section 7. Thus, the theoretical bias for the model parameter and marginal effect are 15% and -0.23%, and their Monte Carlo estimates are 18% and -1% (see Tables 2 and 3).

When can we use uncorrected fixed effects estimators of marginal effects in practice? The expression of the bias derived in Proposition 2 is also useful to answer this question. Thus, since the bias is a fixed effects average, its value can be estimated in the sample using the procedure described in HN. Moreover, a standard Wald test can be constructed to determine whether the

[27]The Nerlove process is not normalized to help me compare with the results of the Monte Carlo in Section 7.

bias is significantly different from zero.

6 Extension: Dynamic Discrete Choice Models

6.1 The Model

Consider now the following dynamic version of the panel discrete choice model

$$Y = 1\left\{\theta_{y,0}Y_{-1} + X'\theta_{x,0} + \alpha - \epsilon \geq 0\right\}, \tag{6.1}$$

where Y_{-1} is a binary random variable that takes on value one if the outcome occurred in the previous period and zero otherwise. The rest of the variables are defined as in the static case. In this model, persistence in the outcome can be a consequence of higher unobserved individual propensity to experience the event in all the periods, as measured by α, or to alterations in the individual behavior for having experienced the event in the previous period, as measured by $\theta_{y,0}Y_{-1}$. Heckman (1981) refers to these sources of persistence as heterogeneity and true state dependence, respectively. Examples of empirical studies that use this type of specification include Card and Sullivan (1988), Moon and Stotsky (1993), Roberts and Tybout (1997), Hyslop (1999), Chay and Hyslop (2000), and Carro (2003).

To estimate the model parameters, I adopt a fixed-effects estimation approach. This approach has the additional advantage in dynamic models of not imposing restrictions on the initial conditions of the process (Heckman, 1981). Then, given a sample of the observable variables and assuming a distribution for ϵ conditional on $(Y_{-1}, \bar{X}, \alpha)$, the model parameters can be estimated by maximum likelihood conditioning on the initial observation of the sample.

6.2 Large-T Approximation to the Bias

In the presence of dynamics, fixed effects MLEs of structural parameters suffer from the incidental parameters problem even when the model is linear; see, for e.g., Nerlove (1967), Nerlove (1971), and Nickell (1981). Formulae for the bias can be obtained using large-T asymptotic expansions of the estimators, which in this case include Hurwicz-type terms due to the correlations between the observations. Thus, let $\bar{Z}_t$ denote $(\bar{X}, Y_{t-1}, ..., Y_0)$, and $E_{T-j}[z_{t-k}]$ denote $\lim_{T\to\infty}\sum_{t=k+1}^{T} E\left[z_{t-k}|\bar{Z}_{t-j}, \alpha\right]/(T-k)$ for $k \leq j$, then a large-T expansion for the estimators of the individual effects can be constructed as

$$\widehat{\alpha}_i = \alpha_i + \psi_i^d/\sqrt{T} + \beta_i^d/T + o_p(1/T), \quad \psi_i^d = \sum_{t=1}^{T} \psi_{it}^d/\sqrt{T} \xrightarrow{d} \mathcal{N}(0, \sigma^{2,d}), \tag{6.2}$$

$$\psi_{it}^d = -E_T\left[v_{it\alpha}\right]^{-1} v_{it}, \quad \sigma_i^{d,2} = -E_T\left[v_{it\alpha}\right]^{-1} + 2 \lim_{T\to\infty} \sum_{j=1}^{T-1} E_{T-j}\left[\psi_{it}^d \psi_{i,t-j}^d\right], \tag{6.3}$$

$$\beta_i^d = -E_T\left[v_{it\alpha}\right]^{-1} \left\{ \lim_{T\to\infty} \sum_{j=0}^{T-1} E_{T-j}\left[v_{it\alpha}\psi_{i,t-j}^d\right] + \frac{1}{2}\sigma_i^2 E_T\left[v_{it\alpha\alpha}\right] \right\}. \tag{6.4}$$

As in the static model, ψ_{it}^d, $\sigma_i^{d,2}$ and β_i^d are the influence function, first-order variance and higher-order bias of $\widehat{\alpha}_i$ as $T \to \infty$. For the common parameter, the expressions for the Jacobian and the first term of the bias of the estimating equation are

$$\mathcal{J}^d = \bar{E}\left\{ E_T\left[u_{it\theta}\right] - E_T\left[u_{it\alpha}\right] \frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]} \right\}, \tag{6.5}$$

$$b^d = \bar{E}\left\{ E_T\left[u_{it\alpha}\right]\beta_i^d + \lim_{T\to\infty} \sum_{j=0}^{T-1} E_{T-j}\left[u_{it\alpha}\psi_{i,t-j}^d\right] + \frac{1}{2}\sigma_i^{d,2} E_T\left[u_{it\alpha\alpha}\right] \right\}. \tag{6.6}$$

The first term of the bias of the fixed effects estimator of θ_0 is then $\mathcal{B}^d = -\left(\mathcal{J}^d\right)^{-1} b^d$. This expression corresponds to the bias formula for general nonlinear panel models derived in Hahn and Kuersteiner (2003), where all the terms that depend on Y have been replaced for their conditional expectation given $(\bar{X}, Y_{-1}, \alpha)$. This adjustment removes zero conditional mean terms without affecting the asymptotic properties of the correction.[28] Monte Carlo results in Section 7 show that this modification improves the finite sample performance of the correction for a dynamic logit model.

6.3 Marginal Effects

Marginal effects can be defined in the same manner as for the static model. Bias corrections for fixed effects estimators also extend naturally to the dynamic case by adding some correlation terms. Thus, using the notation of Section 5, the estimator of the bias can be formed as

$$\begin{aligned}\tilde{\Delta}^d = & \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{ m_\alpha\left(r_{it},\tilde{\theta},\tilde{\alpha}_i\right)\left[\widehat{\beta}_i^d\left(\tilde{\theta}\right) + \widehat{\psi}_{it}^d\left(\tilde{\theta}\right)\right] + \frac{1}{2} m_{\alpha\alpha}\left(r_{it},\tilde{\theta},\tilde{\alpha}_i\right)\widehat{\sigma}_i^{d,2}\left(\tilde{\theta}\right)\right\} \\ & + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{j=1}^{J} m_\alpha\left(r_{it},\tilde{\theta},\tilde{\alpha}_i\right)\widehat{\psi}_{i,t-j}^d\left(\tilde{\theta}\right)\right\}. \end{aligned} \tag{6.7}$$

[28] Appendix A derives the bias formulae for dynamic discrete choice models, and Appendix B describes the corrections.

Here, J is a bandwidth parameter that needs to be chosen such that $J/T^{1/2} \to 0$ as $T \to \infty$ under suitable mixing conditions, see Hahn and Kuersteiner (2003).

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

$$S = \bar{E}\left\{\Phi\left(\frac{\theta_0\ r + \alpha_i}{\sigma}\right)\right\}, \quad \widehat{S} = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{\widehat{\theta}\ r + \widehat{\alpha}_i(\widehat{\theta})}{\widehat{\sigma}}\right), \tag{6.8}$$

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

$$\widehat{\alpha}_i(\theta_T) = \alpha_i + v_i - (\theta_T - \theta_0)\frac{\alpha_i}{1-\theta_0} + o_p(1/T), \quad v_i \sim \mathcal{N}(0, \sigma^2/T). \tag{6.9}$$

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

$$\begin{aligned}
\widehat{S} \quad &\xrightarrow{p} \quad \bar{E}\left\{E\left[\Phi\left(\frac{\theta_T\ u + \widehat{\alpha}_i(\theta_T)}{\sigma_T}\right)\right]\right\} \\
&= \quad \bar{E}\left\{E\left[\Phi\left(\frac{\theta_0 u + \alpha_i + v_i + (\theta_T - \theta_0)\left(u - \frac{\alpha_i}{1-\theta_0}\right) + o_p(T^{-1})}{\sigma_T}\right)\right]\right\} \\
&= \quad \bar{E}\left\{\Phi\left(\frac{\theta_0 u + \alpha_i - \frac{1}{T}(1+\theta_0)\left(u - \frac{\alpha_i}{1-\theta_0}\right) + o(T^{-1})}{\sigma}\right)\right\},
\end{aligned} \tag{6.10}$$

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

In (6.10) we can see that averaging across individuals eliminates the bias of $\widehat{\sigma}^2$, but does not affect the bias of $\widehat{\theta}$. The sign of the bias of $\widehat{S}$ generally depends on the distribution of the individual effects. When there is no heterogeneity, for example, $\widehat{S}$ underestimates (overestimates) the underlying survivor probability when evaluated at values above (below) the unconditional mean of the response, $\alpha/(1-\theta_0)$. This means that if the marginal effects are thought of as differences in survivor probabilities evaluated at two different values u_1 and u_0, fixed effects estimates of marginal effects would be biased downward if the values chosen are $u_1 = 1$ and $u_0 = 0$. For exogenous variables, Monte Carlo results suggest that the bias problem is less severe (see Section 7). Intuitively, it seems that the part of the bias due to the dynamics is less important for the exogenous regressors.[29]

7 Monte Carlo Experiments

This section reports evidence on the finite sample behavior of fixed effects estimators of model parameters and marginal effects for static and dynamic models. In particular, I analyze the finite sample properties of uncorrected and bias-corrected fixed effects estimators in terms of bias and inference accuracy of the asymptotic distribution. The small bias property for marginal effects is illustrated for several lengths of the panel. Robustness of the estimators to small deviations from correct specification is also considered. Thus, the performance of probit and logit estimators is evaluated when the error term is logistic and normal, respectively. All the results presented are based on 1000 replications, and the designs are as in Heckman (1981), Greene (2002), and HN for the static probit model, and as in Honoré and Kyriazidou (2000), Carro (2003), and Hahn and Kuersteiner (2003) for the dynamic logit model.

7.1 Static Model

The model design is

$$y_{it} = 1\left\{x_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\right\}, \ \epsilon_{it} \sim \mathcal{N}(0,1), \ \alpha_i \sim \mathcal{N}(0,1), \tag{7.1}$$

$$x_{it} = t/10 + x_{i,t-1}/2 + u_{it}, \quad x_{i0} = u_{i0}, \quad u_{it} \sim \mathcal{U}(-1/2, 1/2), \tag{7.2}$$

$$n = 100, \ T = 4, \ 8, \ 12; \ \theta_0 = 1, \tag{7.3}$$

where $\mathcal{N}$ and $\mathcal{U}$ denote normal and uniform distribution, respectively. Throughout the tables reported, SD is the standard deviation of the estimator; $\widehat{p}; \#$ denotes a rejection frequency with

[29] For example, assume that in the previous dynamic linear model we add an exogenous regressor X with coefficient β_0, such that $x_{it} \sim$ *i.i.d.* $(0, \sigma_x^2)$ (the individual means are absorbed in the individual effect). Then, it can be shown that the fixed effects estimator of β_0 is fixed-T consistent, and the estimators of θ_0 and σ^2 have the same probability limits as before. The fixed effect estimator of the survivor probability, evaluated at (u_y, u_x), converges to, as $n, T \to \infty$,

$$\widehat{S} = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{\widehat{\theta} u_y + \widehat{\beta} u_x + \widehat{\alpha}_i}{\widehat{\sigma}}\right) \xrightarrow{p} \bar{E}\left\{\Phi\left(\frac{\theta_0 u_y + \beta_0 u_x + \alpha_i - \frac{1}{T}(1+\theta_0)\left(u_y - \frac{\alpha_i}{1-\theta_0}\right) + o(T^{-1})}{\sigma}\right)\right\}. \tag{6.11}$$

Hence, if there is no individual heterogeneity and the probability is evaluated at the unconditional mean of the lagged endogenous variable, i.e. $u_y = \alpha/(1-\theta_0)$, then the fixed effects estimator of the survivor probability is large-T-consistent for every value of u_x. As a result, the derivative with respect to u_x, which is the analog of the marginal effect of X, is also large-T-consistent.

specifying the nominal value; SE/SD is the ratio of the average standard error to standard deviation; and MAE denotes median absolute error.[30] $BC1$ and $BC2$ correspond to the one-step analytical bias-corrected estimators of HN based on maximum likelihood setting and general estimating equations, respectively. JK is the bias correction based on the leave-one-period-out Jackknife-type estimator, see HN. $BC3$ is the one-step bias-corrected estimator proposed here. $CLOGIT$ denotes Andersen's (1973) conditional logit estimator, which is fixed-T consistent when the disturbances are logistically distributed. Iterated and score-corrected estimators are not considered because they are much more cumbersome computationally.[31]

Table 2 gives the Monte Carlo results for the estimators of θ_0 when ϵ_{it} is normally distributed. Both probit and logit estimators are considered and logit estimates are normalized to help compare to the probit.[32] The results here are similar to previous studies (Greene, 2002; and HN) and show that the probit MLE is severely biased, even when $T = 12$, and has important distortions in rejection probabilities. $BC3$ has negligible bias, relative to standard deviation, and improves in terms of bias and rejection probabilities over HN's analytical and jackknife bias-corrected estimators for small sample sizes. It is also remarkable that all the bias corrections and the conditional logit estimator are robust to the type of misspecification considered, even for a sample size as small as $T = 4$. This resembles the well-known similarity between probit and logit estimates in cross sectional data, see Amemiya (1981), but it is more surprising here since the bias correction formulae and conditional logit estimator rely heavily on the form of the true likelihood.

Table 3 reports the ratio of estimators to the truth for marginal effects. Here, I include also two estimators of the average effect based on linear probability models. $LPM - FS$ is the standard linear probability model that uses all the observations; LPM is an adjusted version that calculates the slope from individuals that change status during the sample, i.e., excluding individuals with $y_{it} = 1\ \forall t$ or $y_{it} = 0\ \forall t$, and assigns zero effect to the rest. The results are similar to HN and show small bias in uncorrected fixed effects estimators of marginal effects. Rejection frequencies are higher than their nominal levels, due to underestimation of dispersion. As in cross-section models (Angrist, 2001), both linear models work fairly well in estimating the average effect.[33]

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

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

[32] Estimates and standard errors for logit are multiplied by $\sqrt{3}/\pi$ in order to have the same scale as for probit.

[33] Stoker (1986) shows that linear probability models estimate consistently average effects in index models (e.g., probit and logit) under normality of regressors and individual effects. In general, however, the bias of the linear

7.2 Dynamic Model

The model design is

$$y_{i0} = 1\left\{\theta_{X,0}x_{i0} + \alpha_i - \epsilon_{i0} \geq 0\right\}, \tag{7.4}$$

$$y_{it} = 1\left\{\theta_{Y,0}y_{i,t-1} + \theta_{X,0}x_{it} + \alpha_i - \epsilon_{it} \geq 0\right\}, \quad t = 1, ..., T-1, \tag{7.5}$$

$$\epsilon_{it} \sim \mathcal{L}(0, \pi^2/3), \quad x_{it} \sim \mathcal{N}\left(0, \pi^2/3\right), \tag{7.6}$$

$$n = 250;\ T = 8,\ 12,\ 16;\ \theta_{Y,0} = .5;\ \theta_{X,0} = 1, \tag{7.7}$$

where $\mathcal{L}$ denotes the standardized logistic distribution. Here, the individual effects are correlated with the regressor. In particular, to facilitate the comparison with other studies, I follow Honoré and Kyriazidou (2000) and generate $\alpha_i = \sum_{t=0}^{3} x_{it}/4$. The measures reported are the same as for the static case, and logit and probit estimators are considered.[34] $BC1$ denotes the bias-corrected estimator of Hahn and Kuersteiner (2003); HK is the dynamic version of the conditional logit of Honoré and Kyriazidou (2000), which is fixed-T consistent; MML is the Modified MLE for dynamic models of Carro (2003); and $BC3$ is the bias-corrected estimator that uses conditional expectations in the derivation of the bias formulae.[35] For the number of lags, I choose a bandwidth parameter $J = 1$, as in Hahn and Kuersteiner (2003).

Tables 4 and 5 present the Monte Carlo results for the structural parameters $\theta_{Y,0}$ and $\theta_{X,0}$. Overall, all the bias-corrected estimators have smaller finite sample bias and better inference properties than the uncorrected MLEs. Large-T consistent estimators have median absolute error comparable to HK for $T = 8$.[36] Among bias-corrected estimators, $BC3$ and MML are slightly superior to $BC1$, but there is no clear ranking between them. Thus, $BC3$ has smaller bias and MAE for $\theta_{Y,0}$, but has larger bias and MAE for $\theta_{X,0}$. As for the static model, the bias corrections are robust to the type of misspecification considered for moderate T.

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

probability model depends on the covariance between the conditional probability of the index and the deviations from normality of the index, and on the covariance between the conditional probability of the index and the deviations from linearity of the conditional expectation of regressors and individual effects given the index (see equation (6.1) in Stoker, 1986).

[34]Probit estimates and standard errors are multiplied by $\pi/\sqrt{3}$ to have the same scale as for logit.

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

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

effects are calculated using expression (5.5) with $x_1 = 0$ for the lagged dependent variable, and expression (5.6) for the exogenous regressor. Here, I present results for MLE, $BC1$, $BC3$, linear probability models (LPM and $LPM-FS$), and bias-corrected linear models ($BC-LPM$ and $BC-LPM-FS$) constructed using Nickell's (1981) bias formulae. As in the example of the linear model in Section 6, uncorrected estimates of the effects of the lagged dependent variable are biased downward. Uncorrected estimates of the effect for the exogenous variable, however, have small biases. Large-T corrections are effective in reducing bias and fixing rejection probabilities for both linear and nonlinear estimators.

8 Empirical Application: Female Labor Force Participation

The relationship between fertility and female labor force participation is of longstanding interest in labor economics and demography. For a recent discussion and references to the literature, see Angrist and Evans (1998). Research on the causal effect of fertility on labor force participation is complicated because both variables are jointly determined. In other words, there exist multiple unobserved factors (to the econometrician) that affect both decisions. Here, I adopt an empirical strategy that aims to solve this omitted variables problem by controlling for unobserved individual time-invariant characteristics using panel data. Other studies that follow a similar approach include Heckman and MaCurdy (1980), Heckman and MaCurdy (1982), Hyslop (1999), Chay and Hyslop (2000), Carrasco (2001), and Carro (2003).

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

$$P_{it} = \mathbf{1}\left\{\delta_t + P_{i,t-1}\theta_P + X_{it}'\theta_X + \alpha_i - \epsilon_{it} \geq 0\right\}, \tag{8.1}$$

where P_{it} is the labor force participation indicator; δ_t is a period-specific intercept; $P_{i,t-1}$ is the participation indicator of the previous period; and X_{it} is a vector of time-variant covariates that includes three fertility variables - the numbers of children aged 0-2, 3-5, and 6-17 -, log of husband's earnings, and a quadratic function of age.[37]

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

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

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

Descriptive statistics for the sample are shown in Table 8. Twenty-one percent of the sample is black, and the average age in 1985 was 37. Roughly 72% of women participate in the labor force at some period, the average schooling is 12 years, and the average numbers of children are .2, .3 and 1.1 for the three categories 0-2 year-old, 3-5 year-old, and 6-17 year-old children, respectively.[38] Women that change participation status during the sample, in addition to be younger, less likely to be black, and less educated, have more dependent children and their husband's earnings are slightly higher than average. Interestingly, women who never participate do not have more children than women who are employed each year, though this can be explained in part by the non-participants being older. All the covariates included in the empirical specification display time variation over the period considered.

Table 9 reports fixed effects estimates of index coefficients and marginal effects obtained from a static specification, that is, excluding the lag of participation in equation (8.1). Estimators are labeled as in the Monte Carlo example. The results show that uncorrected estimates of index coefficients are about 15 percent larger than their bias-corrected counterparts; whereas the corresponding differences for marginal effects are less than 2 percent, and insignificant relative to standard errors. It is also remarkable that all the corrections considered give very similar estimates for both index coefficients and marginal effects (for example, bias-corrected logit estimates are the same as conditional logit estimates, up to two decimal points).[39] The adjusted linear probability model gives estimates of the marginal effects closer to logit and probit than the standard linear model. According to the static model estimates, an additional child aged less than 2 reduces the probability of participation by 9 percent, while each child aged 3-5 and 6-17 reduces the probability of participation by 5 percent and 2 percent, respectively.

In the presence of positive state dependence, estimates from a static model overstate the effect of fertility because additional children reduce the probability of participation and participation is positively serially correlated. This can be seen in Table 10, which reports fixed effects estimates of index coefficients and marginal effects using a dynamic specification. Here, as in the Monte Carlo example, uncorrected estimates of the index and effect of the lagged dependent variable

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

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

are significantly smaller (relative to standard errors) than their bias-corrected counterparts for both linear and nonlinear models. Moreover, unlike in the Monte Carlo examples, uncorrected estimates of the effects of the regressors are biased away from zero. Bias-corrected probit gives estimates of index coefficients very similar to probit Modified Maximum Likelihood.[40] The adjusted linear probability model, again, gives estimates of the average effects closer to logit and probit than the standard linear model. Each child aged 0-2 and 3-5 reduces the probability of participation by 6 percent and 3 percent, respectively; while an additional child aged more than 6 years does not have a significant effect on the probability of participation (at the 5 percent level). Finally, a one percent increase in the income earned by the husband reduces a woman's probability of participation by about 0.03%. This elasticity is not sensitive to the omission of dynamics or to the bias corrections.

9 Summary and conclusions

This paper derives bias-corrected fixed effects estimators for model parameters of panel discrete choice models that have better asymptotic and finite sample properties than other similar corrections. The idea behind these corrections is analogous to the use of the conditional information matrix in the variance estimation problem. Thus, the corrections presented here are based on bias formulae that use more intensively the parametric structure of the problem by taking conditional expectations given regressors and individual effects.

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

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

[40] Modified Maximum Likelihood estimates are taken from Carro (2003)

A Bias Formulae for Binary Choice Models

A.1 Static Case

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

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

$$v_{it}(\theta,\alpha_i) = H_{it}(\theta,\alpha_i)\left(y_{it}-F_{it}(\theta,\alpha_i)\right), \quad u_{it}(\theta,\alpha_i) = v_{it}(\theta,\alpha_i)x_{it}, \quad (A.2)$$

where $F_{it}(\theta,\alpha_i)$ denotes $F_\epsilon(x_{it}'\theta+\alpha_i|\bar{X}=\bar{x}_i,\alpha=\alpha_i)$, $H_{it}(\theta,\alpha_i) = f_{it}(\theta,\alpha_i)/\left[F_{it}(\theta,\alpha_i)\left(1-F_{it}(\theta,\alpha_i)\right)\right]$, and f is the pdf associated to F.

Next, since by the Law of Iterated Expectations $E_Z\left[h(z_{it})|\alpha\right] = E_X\left[E_Y\left[h(z_{it})|X,\alpha\right]|\alpha\right]$ for any function $h(z_{it})$, taking conditional expectations of the expressions for the components of the bias in Section 2 yields

$$\sigma_i^2 = E_T\left[H_{it}f_{it}\right]^{-1}, \quad \beta_i = -\sigma_i^4 E_T\left[H_{it}g_{it}\right]/2, \quad (A.3)$$

$$b = -\bar{E}\left\{E_T\left[H_{it}f_{it}x_{it}\right]\beta_i + E_T\left[H_{it}g_{it}x_{it}\right]\sigma_i^2/2\right\}, \quad (A.4)$$

$$\mathcal{J} = -\bar{E}\left\{E_T\left[H_{it}f_{it}x_{it}x_{it}'\right] - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]E_T\left[H_{it}f_{it}x_{it}'\right]\right\}, \quad (A.5)$$

where g denotes the derivative of f and all the expressions are evaluated at the true parameter value (θ_0,α_i).

A.2 Dynamic Case

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

$$l_{it}(\theta,\alpha_i) = y_{it}\log F_{it}(\theta,\alpha_i) + (1-y_{it})\log(1-F_{it}(\theta,\alpha_i)), \quad (A.6)$$

$$v_{it}(\theta,\alpha_i) = H_{it}(\theta,\alpha_i)\left(y_{it}-F_{it}(\theta,\alpha_i)\right), \quad u_{it}(\theta,\alpha_i) = v_{it}(\theta,\alpha_i)x_{it}, \quad (A.7)$$

where $F_{it}(\theta,\alpha_i)$ denotes $F_\epsilon(\theta_y y_{i,t-1}+x_{it}'\theta_x+\alpha_i|Y_{t-1}=y_{i,t-1},...,Y_0=y_0,\bar{X}=\bar{x}_i,\alpha=\alpha_i)$, $H_{it}(\theta,\alpha_i) = f_{it}(\theta,\alpha_i)/\left[F_{it}(\theta,\alpha_i)\left(1-F_{it}(\theta,\alpha_i)\right)\right]$, and f is the pdf associated to F.

Next, taking conditional expectations of the expressions for the components of the bias in Section 6 and using the formulae for the static case, yields

$$\psi_{it}^d = \sigma_i^2 v_{it}, \quad \sigma_i^{2,d} = \sigma_i^2 + 2\lim_{T\to\infty}\sum_{j=1}^{T-1} E_{T-j}\left[\psi_{it}^d\psi_{i,t-j}^d\right], \quad (A.8)$$

$$\beta_i^d = \beta_i + \sigma_i^2\lim_{T\to\infty}\sum_{j=1}^{T-1}\left\{E_{T-j}\left[H_{it}f_{it}\psi_{i,t-j}^d\right] + E_{T-j}\left[\psi_{it}^d\psi_{i,t-j}^d\right]E_T\left[H_{it}g_{it}+2G_{it}f_{it}\right]\right\}, \quad (A.9)$$

$$b^d = -\bar{E}\left\{E_T\left[H_{it}f_{it}x_{it}\right]\beta_i^d + E_T\left[H_{it}g_{it}x_{it}\right]\sigma_i^2/2 + \lim_{T\to\infty}\sum_{j=1}^{T-1}E_{T-j}\left[H_{it}f_{it}\psi_{i,t-j}^d x_{it}\right]\right\} \quad (A.10)$$

$$- \bar{E}\left\{\lim_{T\to\infty}\sum_{j=1}^{T-1}E_{T-j}\left[\psi_{it}^d\psi_{i,t-j}^d\right]E_T\left[H_{it}g_{it}x_{it}+2G_{it}f_{it}x_{it}\right]\right\}, \quad (A.11)$$

$$\mathcal{J}^d = \mathcal{J}, \quad (A.12)$$

where g denotes the derivative of f, $G_{it} = \left(g_{it}F_{it}(1-F_{it}) - f_{it}^2(1-2F_{it})\right) / \left[F_{it}(1-F_{it})\right]^2$ is the derivative of H_{it}, and all the expressions are evaluated at the true parameter value (θ_0, α_i).

B Bias Corrections in Dynamic models

Here, I use the expressions in Appendix A to construct bias-corrected estimators for the dynamic model. Let $\theta \equiv (\theta_y, \theta_x')'$, $z_{it} \equiv (y_{i,t-1}, x_{it}')'$ and

$$\begin{array}{ll} F_{it}(\theta) \equiv F_\epsilon(z_{it}'\theta + \widehat{\alpha}_i(\theta)|\bar{Z}_t, \alpha), & f_{it}(\theta) \equiv f_\epsilon(z_{it}'\theta + \widehat{\alpha}_i(\theta)|\bar{Z}_t, \alpha), \\ g_{it}(\theta) \equiv f_\epsilon'(z_{it}'\theta + \widehat{\alpha}_i(\theta)|\bar{Z}_t, \alpha), & H_{it}(\theta) \equiv \frac{f_{it}(\theta)}{F_{it}(\theta)(1-F_{it}(\theta))}, \\ G_{it}(\theta) \equiv \frac{g_{it}(\theta)}{F_{it}(\theta)(1-F_{it}(\theta))} - \frac{f_{it}(\theta)^2(1-2F_{it}(\theta))}{[F_{it}(\theta)(1-F_{it}(\theta))]^2}. & \end{array} \tag{B.1}$$

Then, the components of the large-T expansion for the estimator of the individual effects can be estimated adding some terms to the analogous expressions for the static case. Thus,

$$\widehat{\psi}_{it}^d(\theta) = H_{it}(\theta)\left[y_{it} - z_{it}'\theta - \widehat{\alpha}_i(\theta)\right], \tag{B.2}$$

$$\widehat{\sigma}_i^{d,2}(\theta) = \widehat{\sigma}_i^2(\theta) + 2\sum_{j=1}^{J} \widehat{E}_{T-j}\left[\widehat{\psi}_{it}^d(\theta)\widehat{\psi}_{i,t-j}^d(\theta)\right], \tag{B.3}$$

$$\begin{aligned} \widehat{\beta}_i^d(\theta) &= \widehat{\beta}_i(\theta) + \widehat{\sigma}_i^2(\theta)\sum_{j=1}^{J} \widehat{E}_{T-j}\left[H_{it}(\theta) f_{it}(\theta)\widehat{\psi}_{i,t-j}^d(\theta)\right] \\ &+ \widehat{\sigma}_i^2(\theta)\sum_{j=1}^{J} \widehat{E}_{T-j}\left[\widehat{\psi}_{i,t}^d(\theta)\widehat{\psi}_{i,t-j}^d(\theta)\right] \widehat{E}_T\left[H_{it}(\theta) g_{it}(\theta) + 2G_{it}(\theta) f_{it}(\theta)\right]. \end{aligned} \tag{B.4}$$

Similarly, for the estimator of the common parameter

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

$$\begin{aligned} \widehat{b}^d(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{E}_T\left[H_{it}(\theta)f_{it}(\theta)x_{it}\right]\widehat{\beta}_i^d(\theta) + \sum_{j=1}^{J}\widehat{E}_{T-j}\left[H_{it}(\theta)g_{it}(\theta)x_{it}\widehat{\psi}_{i,t-j}^d(\theta)\right]\right\} \\ &- \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{J}\widehat{E}_{T-j}\left[\widehat{\psi}_{i,t}(\theta)\widehat{\psi}_{i,t-j}(\theta)\right]\widehat{E}_T\left[H_{it}(\theta) g_{it}(\theta)x_{it} + 2G_{it}(\theta) f_{it}(\theta)x_{it}\right]\right\} \\ &- \frac{1}{n}\sum_{i=1}^{n}\widehat{\sigma}^2(\theta)\widehat{E}_T\left[H_{it}(\theta)g_{it}(\theta)x_{it}\right]/2. \end{aligned} \tag{B.6}$$

Here, J is a bandwidth parameter that needs to be chosen such that $J/T^{1/2} \to 0$ as $T \to \infty$ under suitable mixing conditions, see Hahn and Kuersteiner (2003). For the first-order variance of the estimator of the individual effects, a kernel function can be used to guarantee that the estimates are positive, e.g., Newey and West (1987). From these formulae, all the bias-corrected estimators described in Section 3 can be formed.

For dynamic binary choice models, Lancaster (2000) and Woutersen (2002) derive score bias correction methods, and Carro (2003) extends the Modified Maximum Likelihood estimator of Cox and Reid (1987). The exact relationship between all these methods and the approach followed in Hahn and Kuersteiner

(2003) is not yet known. The reason is that the equivalence results from the static model do not generalize directly to the dynamic case.

Finally, note that the asymptotic variance of the estimator of the common parameter needs to be adjusted to take into account the dependence across the observations, and is no longer the inverse of the Jacobian of the estimating equation. In this case, we have the standard sandwich formula

$$\mathcal{V} = \left(-\mathcal{J}^d\right)^{-1} \Omega \left(-\mathcal{J}^d\right)^{-1}, \quad \Omega = \bar{V}\left\{T^{1/2} E_T\left[U_{it}\right]\right\}, \quad U_{it} = u_{it} - v_{it}\frac{E_T\left[u_{it\alpha}\right]}{E_T\left[v_{it\alpha}\right]}, \tag{B.7}$$

where Ω can be estimated using a kernel function to guarantee positive definiteness, see Hahn and Kuersteiner (2003).

C Proofs

C.1 Lemmas

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

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

and

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

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

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

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

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

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

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

C.2 Proof of Proposition 1

Proof. First, note that for the probit $g_{it} = -(x_{it}'\theta_0 + \alpha_i) f_{it}$. Then, substituting this expression for g_{it} in the bias formulae of the static model, see Appendix A, yields

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

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

$$b = -\bar{E}\left\{\sigma_i^4 E_T\left[H_{it} f_{it} x_{it}\right] E_T\left[H_{it} f_{it} x_{it}'\right] \theta_0 + \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right] \alpha_i\right\} / 2 \tag{C.7}$$

$$+ \ \bar{E}\left\{\sigma_i^2 E_T\left[H_{it} f_{it} x_{it} x_{it}'\right] \theta_0 + \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right] \alpha_i\right\} / 2 \tag{C.8}$$

$$= \ \bar{E}\left\{\sigma_i^2 \left(E_T\left[H_{it} f_{it} x_{it} x_{it}'\right] - \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right] E_T\left[H_{it} f_{it} x_{it}'\right]\right)\right\} \theta_0 / 2 = -\bar{E}\left[\sigma_i^2 \mathcal{J}_i\right] \theta_0 / 2. \tag{C.9}$$

Finally, we have for the bias

$$\mathcal{B} = -\mathcal{J}^{-1}b = \frac{1}{2}\bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\theta_0. \tag{C.10}$$

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

C.3 Proof of Proposition 2

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

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

$$\widehat{\mu} \xrightarrow{p} \bar{E}\left\{\theta_T E_T\left[\phi\left(x_{it}'\theta_T + \widehat{\alpha}_i(\theta_T)\right)\right]\right\}. \tag{C.11}$$

Next, we have the following expansion for the limit index, $\widehat{\xi}_{it}(\theta_T) \equiv x_{it}'\theta_T + \widehat{\alpha}_i(\theta_T)$, around θ_0

$$\widehat{\xi}_{it}(\theta_T) \quad = \quad x_{it}'\theta_0 + \widehat{\alpha}_i(\theta_0) + \left[x_{it}' + \frac{\partial\widehat{\alpha}_i(\bar{\theta})}{\partial\theta'}\right](\theta_T - \theta_0). \tag{C.12}$$

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

$$\widehat{\alpha}_i(\theta_0) = \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + R_{1i}/T^{3/2}, \quad \psi_i \xrightarrow{d} \mathcal{N}(0, \sigma_i^2 \equiv -E_T\left[v_{it\alpha}\right]^{-1}), \tag{C.13}$$

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

$$\frac{\partial\widehat{\alpha}_i(\bar{\theta})}{\partial\theta'} = -\frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]} + R_{2i}/\sqrt{T} = \sigma_i^2E_T\left[v_{it\theta}\right] + R_{2i}/\sqrt{T}, \tag{C.14}$$

where $R_{2i} = O_p(1)$ and $E_T\left[R_{2i}/T\right] = O(1)$ uniformly in i, again by the conditions of the proposition. Plugging (C.13) and (C.14) into the expansion for the index in (C.12) yields, for $\xi_{it} = x_{it}'\theta_0 + \alpha_i$,

$$\widehat{\xi}_{it}(\theta_T) \quad = \quad \xi_{it} + \psi_i/\sqrt{T} + \beta_{\xi_i}/T + R_{3i}/T^{3/2}, \tag{C.15}$$

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

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

$$\begin{aligned}
\beta_{\xi_i} \quad &= \quad \left(\sigma_i^4E_T\left[H_{it}f_{it}x_{it}'\right]\theta_0 + \sigma_i^2\alpha_i\right)/2 + \left(x_{it}' - \sigma_i^2E_T\left[H_{it}f_{it}x_{it}'\right]\right)\mathcal{B} + O(T^{-2}) \\
&= \quad \sigma_i^2\xi_{it}/2 - \left(x_{it}' - \sigma_i^2E_T\left[H_{it}f_{it}x_{it}'\right]\right)\sigma_i^2\theta_0/2 + \left(x_{it}' - \sigma_i^2E_T\left[H_{it}f_{it}x_{it}'\right]\right)\bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\theta_0 + O(T^{-2}) \\
&= \quad \sigma_i^2\xi_{it}/2 - \mathcal{D}_i + O(T^{-2}),
\end{aligned} \tag{C.16}$$

where $\mathcal{D}_i = \left(x_{it}' - \sigma_i^2E_T\left[H_{it}f_{it}x_{it}'\right]\right)\left(\sigma_i^2\mathcal{I}_p - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\right)\theta_0/2$, and $\mathcal{I}_p$ denotes the $p \times p$ identity matrix. Substituting the expression for β_{ξ_i} in (C.15) gives

$$\widehat{\xi}_{it}(\theta_T) = \left[1 + \sigma_i^2/2T\right]\xi_{it} + \psi_i/\sqrt{T} - \mathcal{D}_i/T + R_i/T^{3/2}, \quad \psi_i/\sqrt{T} \overset{a}{\sim} \mathcal{N}(0, \sigma_i^2/T) \tag{C.17}$$

where $R_i = O_p(1)$ and $E_T\left[R_i/T^2\right] = O(1)$ uniformly in i.

Finally, using Lemma 1 and expanding around θ_0, it follows that

$$
\begin{aligned}
\widehat{\mu} &\xrightarrow{p} \bar{E}\left\{\theta_T E_T\left[\phi\left(\widehat{\xi}_i(\theta_T)\right)\right]\right\} = \bar{E}\left\{\theta_T \int \phi\left(\left[1+\sigma_i^2/2T\right]\xi_{it} + v - \mathcal{D}_i/T\right)\frac{\sqrt{T}}{\sigma_i}\;\phi\left(\frac{\sqrt{T}v}{\sigma_i}\right)dv + O(T^{-2})\right\} \\
&= \bar{E}\left\{\left(1+\sigma_i^2/T\right)^{-1/2}\theta_T\phi\left(\frac{\left[1+\sigma_i^2/2T\right]\xi_{it} - \mathcal{D}_i/T}{\sqrt{1+\sigma_i^2/T}}\right) + O(T^{-2})\right\} \\
&= \mu + \frac{1}{2T}\bar{E}\left\{\phi(\xi_{it})\left(\xi_{it}\theta_0\left(x_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]\right)' - \mathcal{I}_p\right)\left(\sigma_i^2\mathcal{I}_p - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\right)\theta_0\right\} + O(T^{-2}) \\
&= \mu + \frac{1}{T}\mathcal{B}_\mu + O\left(T^{-2}\right), \qquad \text{(C.18)}
\end{aligned}
$$

as T grows, since

$$
\frac{\sigma_{iT}^2}{\sqrt{1+\sigma_i^2/T}} = \left(1+\frac{\sigma_i^2}{2T}\right)\left(1-\frac{\sigma_i^2}{2T} + O(T^{-2})\right) = 1 + O(T^{-2}). \qquad \text{(C.19)}
$$

■

D Relationship between Bias Correction of the Estimating Equation and Modified Maximum Likelihood

The first order condition for the MML estimator is

$$
0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}(\check{\theta}) - \frac{1}{T}\tilde{b}(\check{\theta}), \qquad \text{(D.1)}
$$

where

$$
\tilde{b}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{2}\frac{\widehat{E}_T\left[v_{it\alpha\theta}(\theta) + v_{it\alpha\alpha}(\theta)\partial\widehat{\alpha}_i(\theta)/\partial\theta'\right]}{\widehat{E}_T\left[v_{it\alpha}(\theta)\right]} + \frac{\widehat{E}_T\left[\left(H_{it}(\theta)g_{it}(\theta) + G_{it}(\theta)f_{it}(\theta)\right)\left(x_{it} + \partial\widehat{\alpha}_i(\theta)/\partial\theta'\right)\right]}{\widehat{E}_T\left[H_{it}(\theta)f_{it}(\theta)\right]}\right\}. \qquad \text{(D.2)}
$$

From the first order condition for $\widehat{\alpha}_i(\theta)$, note that as $T \to \infty$

$$
\frac{\partial\widehat{\alpha}_i(\theta)}{\partial\theta} = -\frac{E_T\left[v_{it\theta}(\theta)\right]}{E_T\left[v_{it\alpha}(\theta)\right]} + o_p(1). \qquad \text{(D.3)}
$$

Replacing this expression in (D.1), and using the Bartlett identities $v_{it\theta} = u_{it\alpha}$ and $v_{it\alpha\theta} = u_{it\alpha\alpha}$, as $n, T \to \infty$

$$
\begin{aligned}
\tilde{b}(\theta) &\xrightarrow{p} \bar{E}\left\{\frac{1}{2}\frac{E_T\left[H_{it}(\theta)g_{it}(\theta)\right]}{E_T\left[H_{it}(\theta)f_{it}(\theta)\right]^2}E_T\left[H_{it}(\theta)f_{it}(\theta)x_{it}\right] - \frac{1}{2}\frac{E_T\left[H_{it}(\theta)g_{it}(\theta)x_{it}\right]}{E_T\left[H_{it}(\theta)f_{it}(\theta)\right]}\right\} \qquad \text{(D.4)} \\
&= -\bar{E}\left\{E_T\left[H_{it}(\theta)f_{it}(\theta)x_{it}\right]\beta_i(\theta) + E_T\left[H_{it}(\theta)g_{it}(\theta)x_{it}\right]\sigma_i^2(\theta)/2\right\} = b(\theta). \qquad \text{(D.5)}
\end{aligned}
$$

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," *Review of Business and Economic Statistics* 19(1), 2-16.

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

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

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

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

[9] Brown, B. W., and W. K. Newey (1998), "Efficient Semiparametric Estimation of Expectations," *Econometrica* 66, 453-464.

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

[11] Carro, J. M. (2003), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," CEMFI working paper 0304.

[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. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics* 18, 5-46.

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

[16] 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.

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

[18] 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.

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

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

[21] Cox, D. R., and N. Reid (1987), "Parameter Orthogonality and Approximate Conditional Inference," *Journal of the Royal Statistical Society. Series B* 49, 1-39.

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

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

[24] Ferguson, H., N. Reid, and D. R. Cox (1991), "Estimating Equations from Modified Profile Likelihood, " in V. P. Godambe, ed., *Estimating Functions*, 279-293.

[25] Greene, W. H. (2000), *Econometric Analysis*, Fourth Edition, Prentice Hall, New Jersey.

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

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

[28] 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.

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

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

[31] 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.

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

[33] 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.

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

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

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

[37] Kiefer, J., and J. Wolfowitz (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *The Annals of Mathematical Statistics* 27(4), 887-906.

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

[39] Liang, K. Y. (1987), "Estimating Functions and Approximate Conditional Likelihood," *Biometrika* 74, 695-702.

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

[41] McCullagh, P., and R. Tibshirani (1990), "A Simple Method for the Adjustment of Profile Likelihoods," *Journal of the Royal Statistical Society. Series B* 52, 325-344.

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

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

[44] Nelson, D. B. (1995), "Vector Attenuation Bias in the Classical Errors-in-Variables Model," *Economics Letters* 49, 345-349.

[45] 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.

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

[47] 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.

[48] Newey, W. K., and K. D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55(3), 703-708.

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

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

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

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

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

[54] 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.

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

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

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

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

[59] Wooldridge, J. M. (2003), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity ," *Journal of Applied Econometrics*, forthcoming.

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

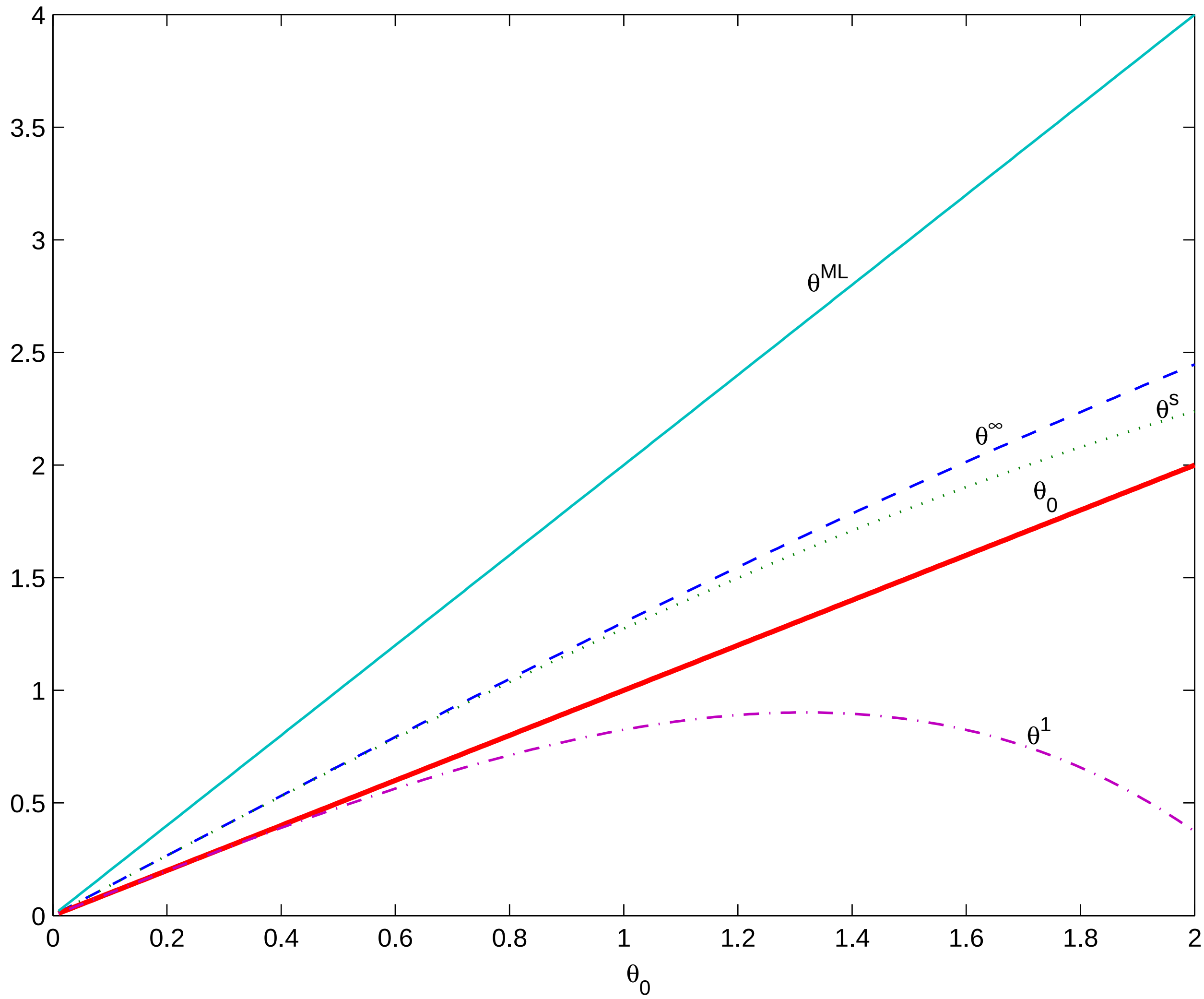

Figure 1: Asymptotic probability limit of estimators: Andersen (1973) two-period logit model. θ_0 is the limit of the conditional logit estimator (true parameter value); θ^{ML} is the limit of the fixed effects maximum likelihood logit estimator; θ^1 is the limit of the one-step bias-corrected estimator; θ^∞ is the limit of the iterated bias-corrected estimator; and θ^s is the limit of the score (estimating equation)-corrected estimator.

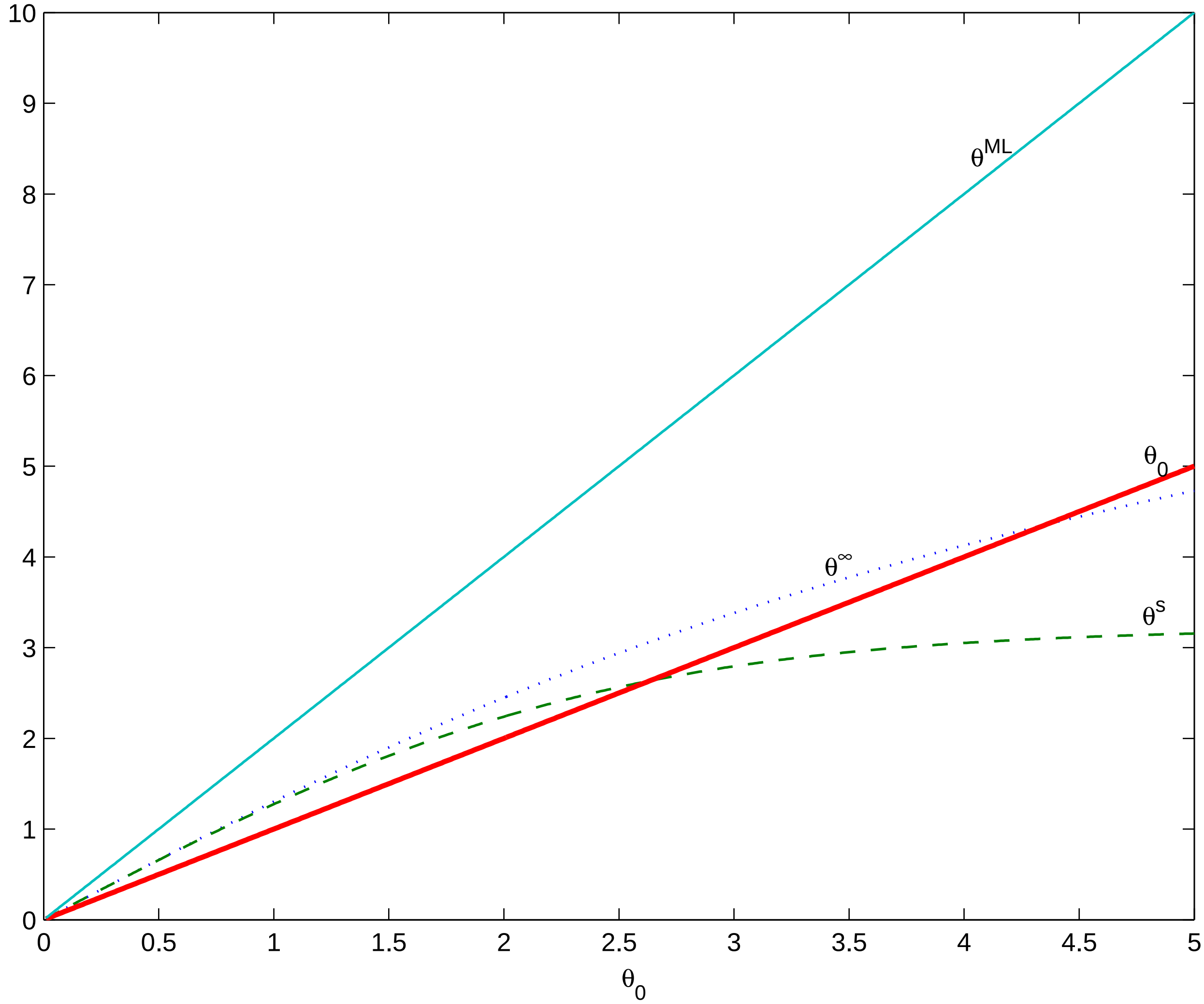

Figure 2: Asymptotic probability limit of estimators: Andersen (1973) two-period logit model. θ_0 is the limit of the conditional logit estimator (true parameter value); θ^{ML} is the limit of the fixed effects maximum likelihood estimator; θ^{∞} is the limit of the iterated bias-corrected estimator; and θ^{s} is the limit of the score (estimating equation)-corrected estimator.

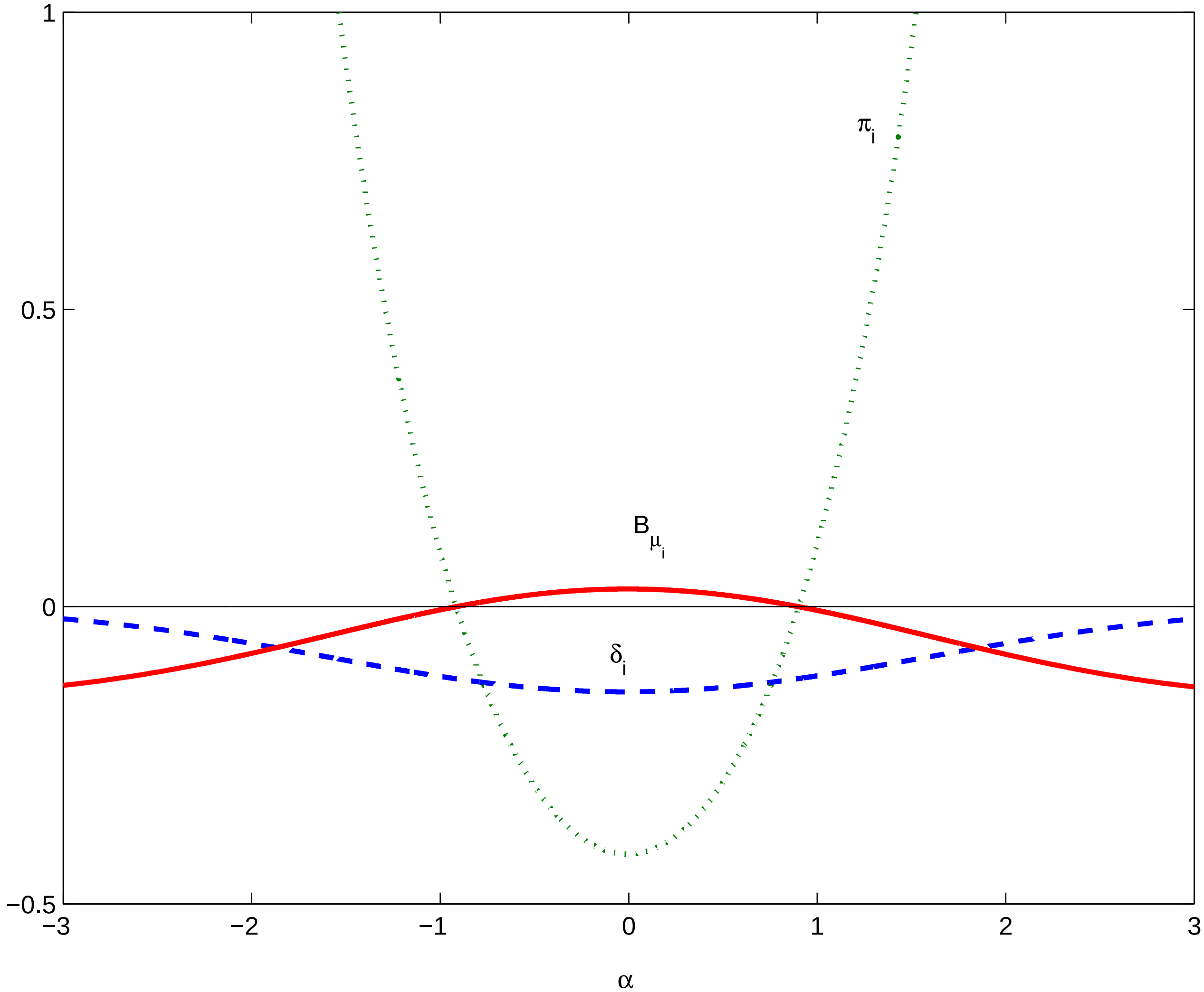

Figure 3: Components of the bias of the fixed effects estimator of the marginal effect: $\mathcal{B}_\mu = E_\alpha[\mathcal{B}_{\mu_i}] = E_\alpha[\delta_i \pi_i]$. Individual effects and regressor generated from independent standard normal distributions.

Table 1: First Order Bias, T = 8
(in percent of the true parameter value)

	Regressor				
Individual Effects	Nerlove	Normal	$\chi^2(1)$	$\chi^2(2)$	Bi(10,.9)
			A - Index Coefficients		
Normal	14.59	15.62	16.56	15.93	15.62
$\chi^2(1)$	12.57	14.77	13.82	14.00	14.85
$\chi^2(2)$	13.17	14.95	13.04	14.55	15.16
Bi(10,.9)	15.46	15.53	16.36	16.47	15.44
			B -Marginal Effects		
Normal	-0.27	-0.07	1.89	0.95	-0.03
$\chi^2(1)$	-0.21	-0.06	-0.18	-0.24	0.34
$\chi^2(2)$	-0.28	-0.05	-0.54	-0.10	0.31
Bi(10,.9)	-0.27	-0.08	1.56	1.59	-0.26

Notes: Bias formulae evaluated numerically using 10,000 replications.

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

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 4				
PROBIT	1.41	1.40	0.393	0.25	0.36	0.82	0.410
JK-PROBIT	0.75	0.75	0.277	0.11	0.19	1.08	0.265
BC1-PROBIT	1.11	1.10	0.304	0.04	0.11	1.03	0.215
BC2-PROBIT	1.20	1.19	0.333	0.09	0.16	0.95	0.253
BC3-PROBIT	1.06	1.06	0.275	0.02	0.06	1.13	0.195
LOGIT	1.30	1.29	0.374	0.17	0.26	0.83	0.331
JK-LOGIT	0.71	0.70	0.239	0.18	0.28	1.15	0.307
BC1-LOGIT	0.97	0.96	0.266	0.04	0.08	1.08	0.180
BC2-LOGIT	0.95	0.94	0.263	0.03	0.09	1.09	0.178
BC3-LOGIT	0.94	0.94	0.253	0.04	0.07	1.13	0.173
CLOGIT	0.95	0.94	0.263	0.04	0.08	1.09	0.177
			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
LOGIT	1.12	1.12	0.148	0.14	0.23	0.91	0.129
JK-LOGIT	0.91	0.91	0.114	0.12	0.20	1.09	0.105
BC1-LOGIT	0.97	0.97	0.127	0.06	0.13	1.00	0.087
BC2-LOGIT	0.96	0.95	0.122	0.07	0.13	1.03	0.087
BC3-LOGIT	0.95	0.95	0.121	0.07	0.13	1.04	0.089
CLOGIT	0.96	0.96	0.122	0.07	0.13	1.03	0.088
			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
LOGIT	1.09	1.09	0.097	0.15	0.25	0.96	0.095
JK-LOGIT	0.95	0.95	0.080	0.08	0.15	1.07	0.068
BC1-LOGIT	0.99	0.99	0.086	0.06	0.10	1.01	0.059
BC2-LOGIT	0.98	0.97	0.083	0.06	0.11	1.04	0.060
BC3-LOGIT	0.98	0.97	0.082	0.06	0.10	1.05	0.059
CLOGIT	0.98	0.98	0.083	0.06	0.10	1.04	0.059

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

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

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 4				
PROBIT	0.99	0.99	0.242	0.10	0.16	0.82	0.163
JK-PROBIT	1.02	1.02	0.285	0.12	0.19	0.75	0.182
BC1-PROBIT	1.00	1.00	0.261	0.12	0.18	0.79	0.176
BC2-PROBIT	1.04	1.04	0.255	0.12	0.19	0.80	0.176
BC3-PROBIT	0.94	0.94	0.226	0.08	0.13	0.91	0.158
LOGIT	1.00	0.99	0.246	0.10	0.16	0.82	0.164
JK-LOGIT	1.01	1.01	0.279	0.13	0.19	0.76	0.189
BC1-LOGIT	0.98	0.97	0.257	0.11	0.17	0.81	0.178
BC2-LOGIT	0.94	0.94	0.236	0.09	0.15	0.88	0.164
BC3-LOGIT	0.93	0.93	0.230	0.08	0.14	0.90	0.166
BC-CLOGIT	0.94	0.94	0.237	0.09	0.15	0.87	0.168
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
LOGIT	0.99	0.99	0.106	0.08	0.13	0.83	0.071
JK-LOGIT	0.99	1.00	0.108	0.07	0.13	0.84	0.072
BC1-LOGIT	1.00	1.00	0.112	0.08	0.15	0.81	0.073
BC2-LOGIT	0.98	0.98	0.106	0.07	0.13	0.85	0.071
BC3-LOGIT	0.98	0.98	0.106	0.08	0.13	0.85	0.071
BC-CLOGIT	0.98	0.98	0.106	0.08	0.13	0.85	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
LOGIT	0.99	0.99	0.063	0.05	0.10	0.77	0.043
JK-LOGIT	1.00	1.00	0.065	0.05	0.11	0.77	0.042
BC1-LOGIT	1.01	1.00	0.067	0.06	0.12	0.74	0.044
BC2-LOGIT	0.99	0.99	0.064	0.06	0.10	0.77	0.043
BC3-LOGIT	0.99	0.99	0.064	0.05	0.10	0.77	0.044
BC-CLOGIT	0.99	0.99	0.064	0.05	0.10	0.77	0.044
LPM	0.99	0.99	0.065	0.06	0.11	0.76	0.041
LPM-FS	1.01	1.01	0.067	0.05	0.11	0.80	0.045

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

Table 4: Estimators of θ_Y ($\theta_{Y,0}$ = .5), $\varepsilon \sim L(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
PROBIT	-0.26	-0.26	0.161	1.00	1.00	0.92	0.763
BC1-PROBIT	0.25	0.25	0.148	0.43	0.54	0.96	0.251
BC3-PROBIT	0.48	0.48	0.140	0.05	0.10	0.99	0.101
LOGIT	-0.24	-0.24	0.154	1.00	1.00	0.92	0.740
BC1-LOGIT	0.39	0.38	0.134	0.14	0.23	1.00	0.131
HK-LOGIT		0.45					0.131
MML-LOGIT		0.39		0.11			0.127
BC3-LOGIT	0.45	0.45	0.134	0.06	0.12	1.00	0.101
			T = 12				
PROBIT	0.06	0.06	0.111	0.97	0.99	0.99	0.435
BC1-PROBIT	0.39	0.39	0.103	0.16	0.26	1.04	0.115
BC3-PROBIT	0.50	0.50	0.101	0.04	0.10	1.05	0.066
LOGIT	0.06	0.06	0.105	0.98	0.99	1.00	0.436
BC1-LOGIT	0.44	0.44	0.094	0.07	0.13	1.07	0.079
BC3-LOGIT	0.47	0.48	0.095	0.05	0.10	1.06	0.064
			T = 16				
PROBIT	0.20	0.20	0.097	0.89	0.93	0.94	0.302
BC1-PROBIT	0.44	0.44	0.091	0.11	0.18	0.98	0.080
BC3-PROBIT	0.51	0.51	0.091	0.06	0.11	0.98	0.061
LOGIT	0.19	0.19	0.091	0.93	0.96	0.95	0.312
BC1-LOGIT	0.46	0.45	0.084	0.07	0.14	1.01	0.067
HK-LOGIT		0.45					0.074
MML-LOGIT		0.48					0.067
BC3-LOGIT	0.48	0.48	0.085	0.06	0.11	1.00	0.059

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

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

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
PROBIT	1.28	1.28	0.082	0.97	0.99	0.92	0.277
BC1-PROBIT	1.19	1.19	0.077	0.78	0.86	0.89	0.187
BC3-PROBIT	1.10	1.09	0.065	0.32	0.44	0.97	0.094
LOGIT	1.22	1.22	0.082	0.84	0.90	0.93	0.222
BC1-LOGIT	1.08	1.08	0.074	0.22	0.32	0.90	0.079
HK-LOGIT		1.01					0.050
MML-LOGIT		1.01		0.06			0.039
BC3-LOGIT	1.05	1.04	0.067	0.12	0.18	0.96	0.054
			T = 12				
PROBIT	1.19	1.19	0.057	0.94	0.97	0.93	0.185
BC1-PROBIT	1.11	1.11	0.053	0.61	0.73	0.92	0.110
BC3-PROBIT	1.08	1.08	0.050	0.35	0.48	0.95	0.076
LOGIT	1.13	1.13	0.057	0.67	0.77	0.94	0.126
BC1-LOGIT	1.03	1.03	0.051	0.10	0.16	0.94	0.039
BC3-LOGIT	1.02	1.02	0.050	0.08	0.13	0.96	0.035
			T = 16				
PROBIT	1.15	1.15	0.045	0.95	0.98	0.97	0.148
BC1-PROBIT	1.09	1.09	0.042	0.59	0.71	0.96	0.090
BC3-PROBIT	1.07	1.07	0.041	0.41	0.54	0.99	0.070
LOGIT	1.09	1.09	0.044	0.53	0.66	0.98	0.088
BC1-LOGIT	1.02	1.01	0.041	0.07	0.13	0.98	0.029
HK-LOGIT		1.01					0.023
MML-LOGIT		1.01					0.029
BC3-LOGIT	1.01	1.01	0.040	0.06	0.10	0.99	0.027

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

Table 6: Estimators of μ_Y (true value = 1), $\varepsilon \sim L(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
PROBIT	-0.38	-0.39	0.235	1.00	1.00	0.92	1.386
BC1-PROBIT	0.43	0.42	0.254	0.68	0.77	0.87	0.577
BC3-PROBIT	0.86	0.85	0.255	0.14	0.22	0.88	0.211
LOGIT	-0.37	-0.37	0.236	1.00	1.00	0.93	1.374
BC1-LOGIT	0.72	0.70	0.257	0.28	0.38	0.88	0.304
BC3-LOGIT	0.85	0.85	0.260	0.14	0.24	0.87	0.212
LPM	-0.40	-0.40	0.238	1.00	1.00	0.92	1.399
BC-LPM	0.77	0.77	0.267	0.23	0.32	0.85	0.261
LPM-FS	-0.45	-0.46	0.257	1.00	1.00	0.92	1.459
BC-LPM-FS	0.84	0.83	0.289	0.16	0.24	0.85	0.236
			T = 12				
PROBIT	0.11	0.11	0.183	1.00	1.00	0.99	0.894
BC1-PROBIT	0.70	0.70	0.189	0.38	0.49	0.96	0.297
BC3-PROBIT	0.93	0.93	0.190	0.08	0.13	0.96	0.136
LOGIT	0.11	0.11	0.182	1.00	1.00	1.00	0.889
BC1-LOGIT	0.86	0.86	0.187	0.13	0.19	0.98	0.169
BC3-LOGIT	0.93	0.94	0.191	0.08	0.13	0.97	0.133
LPM	0.09	0.09	0.188	1.00	1.00	1.00	0.908
BC-LPM	0.91	0.91	0.202	0.09	0.15	0.94	0.151
LPM-FS	0.09	0.09	0.192	1.00	1.00	1.00	0.909
BC-LPM-FS	0.93	0.93	0.207	0.08	0.14	0.94	0.143
			T = 16				
PROBIT	0.34	0.34	0.167	0.98	0.99	0.94	0.660
BC1-PROBIT	0.81	0.80	0.170	0.25	0.35	0.93	0.201
BC3-PROBIT	0.95	0.94	0.171	0.09	0.15	0.92	0.125
LOGIT	0.35	0.34	0.166	0.97	0.99	0.95	0.658
BC1-LOGIT	0.91	0.90	0.168	0.10	0.18	0.95	0.139
BC3-LOGIT	0.95	0.94	0.171	0.08	0.14	0.93	0.128
LPM	0.34	0.33	0.170	0.97	0.99	0.96	0.666
BC-LPM	0.95	0.95	0.180	0.09	0.15	0.91	0.129
LPM-FS	0.34	0.34	0.172	0.97	0.99	0.96	0.665
BC-LPM-FS	0.96	0.95	0.181	0.08	0.15	0.92	0.128

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

Table 7: Estimators of μ_X (true value = 1), $\varepsilon \sim L(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
PROBIT	0.97	0.97	0.041	0.11	0.19	0.93	0.034
BC1-PROBIT	1.02	1.02	0.044	0.11	0.17	0.85	0.032
BC3-PROBIT	0.98	0.97	0.041	0.11	0.19	0.88	0.033
LOGIT	0.98	0.98	0.042	0.07	0.14	0.93	0.031
BC1-LOGIT	1.01	1.01	0.046	0.09	0.15	0.83	0.031
BC3-LOGIT	0.99	0.99	0.043	0.08	0.15	0.86	0.030
LPM	0.96	0.96	0.040	0.19	0.29	0.91	0.043
BC-LPM	0.98	0.97	0.040	0.09	0.16	0.92	0.033
LPM-FS	0.97	0.97	0.040	0.10	0.17	0.97	0.034
BC-LPM-FS	0.99	0.99	0.040	0.05	0.10	0.98	0.028
			T = 12				
PROBIT	0.99	0.99	0.031	0.07	0.14	0.92	0.021
BC1-PROBIT	1.01	1.01	0.032	0.08	0.13	0.89	0.022
BC3-PROBIT	0.99	0.99	0.031	0.07	0.14	0.90	0.020
LOGIT	0.99	0.99	0.031	0.06	0.11	0.93	0.020
BC1-LOGIT	1.01	1.01	0.032	0.06	0.13	0.89	0.022
BC3-LOGIT	1.00	1.00	0.032	0.06	0.13	0.90	0.021
LPM	0.99	0.99	0.031	0.08	0.14	0.93	0.023
BC-LPM	0.99	0.99	0.031	0.06	0.11	0.93	0.021
LPM-FS	0.99	0.99	0.031	0.07	0.12	0.94	0.022
BC-LPM-FS	1.00	1.00	0.031	0.06	0.11	0.95	0.020
			T = 16				
PROBIT	0.99	0.99	0.025	0.05	0.10	0.98	0.017
BC1-PROBIT	1.01	1.01	0.025	0.06	0.11	0.95	0.017
BC3-PROBIT	1.00	1.00	0.025	0.05	0.10	0.96	0.017
LOGIT	1.00	1.00	0.025	0.04	0.09	0.98	0.017
BC1-LOGIT	1.00	1.00	0.026	0.05	0.11	0.95	0.017
BC3-LOGIT	1.00	1.00	0.025	0.05	0.10	0.96	0.017
LPM	0.99	0.99	0.026	0.06	0.12	0.97	0.019
BC-LPM	1.00	1.00	0.026	0.05	0.11	0.98	0.018
LPM-FS	0.99	0.99	0.026	0.05	0.11	0.98	0.019
BC-LPM-FS	1.00	1.00	0.026	0.05	0.10	0.98	0.018

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

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

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

Source: PSID 1980-1988.

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

Estimator	PROBIT			LOGIT				LPM	
	FE	JK	BC3	FE	JK	BC3	C	FE	FE-FS
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
				A - Index Coefficients					
Kids 0-2	-0.71	-0.61	-0.63	-0.68	-0.59	-0.60	-0.60		
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)		
Kids 3-5	-0.42	-0.37	-0.37	-0.40	-0.35	-0.35	-0.35		
	(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	-0.11	-0.11	-0.11		
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		
Log(Husband income)	-0.25	-0.22	-0.22	-0.24	-0.21	-0.21	-0.21		
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
				B - Marginal Effects (%)					
Kids 0-2	-9.22	-9.38	-9.07	-9.35	-9.35	-9.20	-9.20	-9.46	-11.19
	(0.70)	(0.71)	(0.70)	(0.72)	(0.72)	(0.72)	(0.72)	(0.75)	(0.88)
Kids 3-5	-5.45	-5.60	-5.36	-5.53	-5.59	-5.45	-5.45	-5.54	-6.09
	(0.66)	(0.66)	(0.66)	(0.67)	(0.67)	(0.67)	(0.67)	(0.70)	(0.81)
Kids 6-17	-1.68	-1.59	-1.66	-1.78	-1.72	-1.76	-1.76	-1.78	-1.25
	(0.53)	(0.53)	(0.53)	(0.54)	(0.54)	(0.54)	(0.54)	(0.58)	(0.56)
Log(Husband income)	-3.25	-3.31	-3.20	-3.26	-3.29	-3.22	-3.22	-3.17	-3.61
	(0.70)	(0.69)	(0.70)	(0.71)	(0.71)	(0.71)	(0.71)	(0.72)	(0.71)

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

Source: PSID 1980-1988.

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

Estimator	PROBIT FE [1]	PROBIT BC3 [2]	PROBIT MML [3]	LOGIT FE [4]	LOGIT BC3 [5]	LPM FE [6]	LPM BC [7]	LPM FE-FS [8]	LPM BC-FS [9]
				A - Index Coefficients					
Participation$_{t-1}$	0.76	1.04	1.08	0.69	0.96				
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
Kids 0-2	-0.55	-0.44	-0.40	-0.53	-0.43				
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)				
Kids 3-5	-0.29	-0.21	-0.18	-0.27	-0.20				
	(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)	(0.04)	(0.04)	(0.04)	(0.04)				
Log(Husband income)	-0.25	-0.22	-0.21	-0.24	-0.21				
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)				
				B - Marginal Effects (%)					
Participation$_{t-1}$	10.69	16.86		10.47	16.80	11.42	17.80	25.58	39.06
	(0.62)	(0.65)		(0.61)	(0.64)	(0.63)	(0.65)	(1.30)	(1.36)
Kids 0-2	-6.76	-5.95		-6.81	-5.94	-6.87	-5.42	-8.02	-6.35
	(0.75)	(0.72)		(0.75)	(0.72)	(0.75)	(0.73)	(0.86)	(0.84)
Kids 3-5	-3.55	-2.79		-3.53	-2.72	-3.44	-2.27	-3.57	-2.24
	(0.69)	(0.67)		(0.69)	(0.67)	(0.69)	(0.66)	(0.79)	(0.77)
Kids 6-17	-0.91	-0.67		-0.95	-0.69	-0.93	-0.45	-0.49	-0.10
	(0.55)	(0.52)		(0.54)	(0.52)	(0.55)	(0.52)	(0.53)	(0.51)
Log(Husband income)	-3.08	-2.90		-3.07	-2.87	-2.98	-2.88	-3.29	-3.12
	(0.70)	(0.68)		(0.71)	(0.68)	(0.70)	(0.69)	(0.69)	(0.68)

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