The Bias of the Fixed Effects Estimator in Nonlinear Models

William Greene*

Department of Economics, Stern School of Business,
New York University,

October, 2002

Abstract

The nonlinear fixed effects model has two shortcomings, one practical, one methodological. The practical obstacle relates to the difficulty of estimating nonlinear models with possibly thousands of dummy variable coefficients. In fact, in many models of interest to practitioners, estimation of the fixed effects model is feasible even in panels with very large numbers of groups. The result, though not new, appears not to be well known. The more difficult, methodological issue is the incidental parameters problem that raises questions about the statistical properties of the estimator. There is relatively little empirical evidence on the behavior of the fixed effects estimator and that which has been obtained has focused almost exclusively on binary choice models. In this paper, we use Monte Carlo methods to examine the small sample bias in the tobit, truncated regression and Weibull survival models as well as the binary probit and logit and ordered probit discrete choice models. We find that the estimator in the continuous response models behaves quite differently from the familiar and oft cited results. Among our findings are: first, a widely accepted result that suggests that the probit estimator is actually relatively well behaved appears to be incorrect; second, the slopes in the tobit model, unlike the probit and logit models that have been studied previously, are largely unaffected by the incidental parameters problem, but a surprising result related to the disturbance variance estimator arises instead; lest one jump to a conclusion that the finite sample bias is restricted to discrete choice models, we submit evidence on the truncated regression, which is yet unlike the tobit in that regard - it is biased *toward* zero; fourth, we find in the Weibull model that the biases in a vector of coefficients need not be in the same direction;. fifth, as apparently unexamined previously, the estimated asymptotic standard errors for the fixed effects estimators appear uniformly to be downward biased. Finally, we consider directly the issue of "consistency" in the context of the tobit model and find that widely received perceptions to the contrary, at least in this model, the fixed effects estimator appears to be neither biased nor inconsistent. In sum, the finite sample behavior of the fixed effects estimator is much more varied than the received literature would suggest.

Keywords: Panel data, fixed effects, computation, Monte Carlo, tobit, truncated regression, bias, finite sample.

JEL classification: C1, C4

---

* 44 West 4th St., New York, NY 10012, USA, Telephone: 001-212-998-0876; fax: 01-212-995-4218; e-mail: wgreene@stern.nyu.edu, URL www.stern.nyu.edu/~wgreene. This paper has benefited from discussions with George Jakubson, Paul Allison, Martin Spiess, Manuel Arellano and Scott Thompson and from seminar groups at The University of Texas, University of Illinois, Binghamton University, Syracuse University and New York University. Any remaining errors are my own.

1. Introduction

In the analysis of panel data with nonlinear models, researchers often choose between a random effects and a fixed effects specification. The random effects model requires an unpalatable orthogonality assumption - consistency requires that the effects be uncorrelated with the included variables. The fixed effects model relaxes this assumption but the estimator suffers from the 'incidental parameters problem' analyzed by Neyman and Scott (1948) [see, also, Lancaster (2000)]. The fixed effects maximum likelihood estimator is inconsistent when T, the length of the panel is fixed. In the models that have been examined in detail, it appears also to be biased in finite samples. How serious these problems are in practical terms remains to be established - there is only a very small amount of received empirical evidence and very little theoretical foundation. [See, e.g., Maddala (1987) and Baltagi (2000).] Impressions to the contrary notwithstanding, Neyman and Scott did not establish that the fixed effects estimator would generally be biased in a finite sample; they found as a side result in their analysis of *asymptotic* efficiency that the maximum likelihood estimator of the variance in a fixed effects regression model had an exact expectation that was $(T-1)/T$ times the true value. They provided no general results on small T bias. The only received analytic results in this regard are those for the binomial logit model established by Anderson (1973) and Hisao (1996). Other results on this phenomenon are based on Monte Carlo studies of binary choice estimators. [See, e.g., Heckman (1981a) and Katz (2001).]

There is an extensive literature on semiparametric and GMM approaches for some panel data models with latent heterogeneity [see, e.g., Manski (1987), Charlier et al. (1995), Chen et al. (1999), Honoré and Kyriazidou (2000) and Honoré and Lewbel (2002).] Among the practical limitations of these estimators is that although they provide estimators of the primary slope parameters, they usually do not provide estimators for the full set of model parameters and thus preclude computation of marginal effects, probabilities or predictions for the dependent variable. (Indeed, some estimation techniques which estimate only the slope parameters and only "up to scale" provide essentially only information about signs of coefficients and classical ("yes or no") statistical significance of variables in the model..) In contrast, the fixed effects estimator is a full information estimator that, under its assumptions, provides results for all model parameters including the parameters of the heterogeneity. In spite of its shortcomings, the fixed effects estimator has some virtues which suggest that it is worth a detailed look at its properties. This study will examine the behavior of the estimator in a variety of nonlinear models.

Most of the results in the literature are qualitative in nature. One widely cited piece of empirical evidence is Heckman's (1981b) Monte Carlo study of the probit model in which he found that the small sample bias of the estimator appeared to be surprisingly small. However, his study examined a very narrow range of specifications, focused only on the probit model and, did not, in fact, examine a fixed effects model. Heckman analyzed the bias of the fixed effects *estimator* in a random effects *model* – his analysis included the orthogonality assumption noted earlier. In spite of its wide citation, Heckman's results are of limited usefulness for the case in which the researcher contemplates the fixed effects estimator precisely because the assumptions of the random effects model are inappropriate. Moreover, our results below are sharply at odds with Heckman's (even with his specification).

Analysis of the fixed effects model has focused on binary choice models.[1] The now standard result is that the fixed effects estimator is inconsistent and substantially biased *away* from zero when group sizes are small, with a bias that diminishes with increasing group size. We will consider some additional aspects of the estimator. First, the two binary choice estimators that have been examined heretofore are narrow cases. Recent research has been based on an increasing

[1] The model has been studied intensively in the recent literature. A partial list of only the most recent studies of the probit model includes Arellano and Honoré (2001), Cerro (2002), Chen et al. (1999), Hahn (2001), Katz (2001), Laisney and Lechner (2002), Lancaster (1999), and Magnac (2002).

availability of high quality panel data sets and on models that extend well beyond binary choice. There is little received evidence on the behavior of the fixed effects estimator in other models. We will focus on three, the tobit and truncated regression models for limited dependent variables and the Weibull model for survival (duration) data. In the case of the tobit model, a surprising result emerges that would be overlooked by the conventional focus on slope estimators. In brief, the slope estimators in the tobit model appear not to be affected by the incidental parameters problem. But, the problem shows up elsewhere, in the estimated disturbance variance. The truncated regression model behaves quite differently. In this case, both the slopes and the variance are attenuated. No general pattern can be asserted, however. In the Weibull model, two slope coefficients are biased in opposite directions.

This study is organized as follows: We begin in Section 2 with a general specification for nonlinear models with fixed effects. Save for a few well known cases, the potentially huge number of parameters presents a practical problem for estimation of this model. In these few cases, it is possible to condition the constants out of the model, and base estimation of the main parameters on the conditional likelihood. In most cases, this is not possible; for maximum likelihood estimation, all parameters must be estimated simultaneously. Though it appears not to be widely known, in most cases, it is actually possible to estimate the full parameter vector even in models for which there is no conditional likelihood which is free of the nuisance parameters. Some details on computation of the estimator are sketched in Section 2. Section 3 contains two Monte Carlo studies of the fixed effects estimator. We first revisit Heckman's (1981b) study of the probit model as well as the other familiar result, that for the binary logit model. Another discrete choice model that has not been examined previously, the ordered probit model, is examined here as well. An additional question considered in this study has not been addressed previously. Given that the fixed effects estimator is problematic, is it best to ignore the heterogeneity, use a random effects estimator, or use the fixed effects estimator in spite of its shortcomings? The second study considers the tobit and truncated regression models and the Weibull model for censored duration data. Here, we are interested not only in the slope estimators, but the variance estimator and the estimators of marginal effects. We will also examine the estimated standard errors in the fixed effects models. Some conclusions are drawn in Section 4.

The end result of this study is that the fixed effects estimator displays a much greater variety of behavior than suggested in the received literature. Some of the main conclusions of this paper are as follows: First, save for some well documented cases, such as the Poisson model, in which there actually is no incidental parameters problem, the skepticism about the fixed effects estimator is only broadly appropriate. We find that for a wider range of cases for the models that have already been examined in the literature, the estimator is indeed biased, and in a few instances, substantially so even when T is fairly large. Second, Heckman's encouraging results for the probit model appear to be incorrect. However, ignoring heterogeneity (in a probit model) is not necessarily worse than using the fixed effects estimator to account for it. But, using the random effects estimator is worse. Third, the slope estimators in the tobit model do not appear to be affected by the incidental parameters problem. This is an unexpected result, but it must be tempered by a finding that the variance estimator is so affected.   The variance estimator in the tobit model is a crucial parameter for inference and analysis purposes. On the other hand, the bias in the variance estimator appears to fall fairly quickly with increasing T. Even given this additional result, one must look a bit more closely. The marginal effects in the tobit model are much less biased than one might expect. We also find that in cases in which the expected biases in the slope estimators do emerge, it is away from zero, but at the same time, the estimated standard errors appear to be biased toward zero.   The truncated regression model and Weibull models display various patterns that would not be predicted by already received results. Finally, a closer look at the tobit model suggests that in contrast to the widely accepted result, the fixed effects estimator in this one case appears to be consistent as well as unbiased.

2. The Fixed Effects Model and Estimator

We consider a nonlinear model defined by the density for an observed random variable, y_{it},

$$f(y_{it} \mid \mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{i,Ti}) = g(y_{it}, \beta'\mathbf{x}_{it} + \alpha_i, \theta)$$

where β is the vector of slopes, θ is a vector of ancillary parameters such as a disturbance standard deviation, an overdispersion parameter in the Poisson model or the threshold parameters in an ordered probit model. We will leave for future research models with dynamic effects; $y_{i,t-1}$ does not appear on the right hand side of the equation. [See, e.g., Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Orme (1999), Heckman (1978, 1981a), Heckman and MaCurdy (1980), Hahn (2001), Honoré and Kyriazidou (2000)]. The fixed effects model presents two disadvantages. In a few cases, it is possible to condition the possibly large number of constants out of the model, and base estimation of β and θ on a conditional likelihood. But, in most cases, this is not possible; for maximum likelihood estimation, all parameters must be estimated simultaneously. [There are no general results. Lancaster (2000) catalogs those which have been derived.] Though it appears not to be widely known, as discussed below, in most cases, it is actually possible to compute the full parameter vector even in models for which there is no conditional likelihood which is free of the nuisance parameters. Second, because of the incidental parameters problem, the unconditional fixed effects estimator is inconsistent - the asymptotic variance of the estimator of β does not converge to zero as N increases. Moreover, with fixed group sizes, T, there appears to be a significant small sample bias in the estimator. The familiar evidence in this regard is limited to the probit and logit models. (We find, in passing, that the same effect is observed in the ordered probit model.) We will examine the effect further in the context of three models that have continuous dependent variables, the tobit and truncated regression and Weibull duration models. Our results are considerably different from the familiar findings. We will also examine the behavior of the estimator of the asymptotic standard errors for the slope estimators.

2.1. Computation of the Fixed Effects Maximum Likelihood Estimator

The log likelihood function for a sample of *N* repeated observations on group *i* is

$$\log L = \sum_{i=1}^{N}\left[\sum_{t=1}^{T_i}\log g(y_{it},\beta'\mathbf{x}_{it}+\alpha_i,\theta)\right].$$

The likelihood equations for β, θ, and $\alpha = [\alpha_1,...,\alpha_N]'$,

$$\partial \log L / \partial[\beta' \quad \theta' \quad \alpha']' = \mathbf{0},$$

generally do not have explicit solutions for the parameter estimates in terms of the data and must be solved iteratively. In principle, maximization can proceed simply by creating and including a complete set of dummy variables in the model. But, the proliferation of nuisance (incidental) parameters (constant terms) which increase in number with the sample size, ultimately renders conventional gradient based maximization of this likelihood infeasible.

2.2. Conditional Estimation

In the linear case, regression using group mean deviations sweeps out the fixed effects. The *K* slope parameters are estimated by within group least squares, a computation of order *K*, not *N*. A few analogous cases of nonlinear models have been developed, such as the binomial logit model,

$$g(y_{it}, \beta'\mathbf{x}_{it} + \alpha_i) = \Lambda[(2y_{it} - 1)(\beta'\mathbf{x}_{it} + \alpha_i)]$$

where $\Lambda(z) = \exp(z)/[1+\exp(z)]$. [See Chamberlain (1980), Rasch (1960), Krailo and Pike (1984), and Greene (2003, Chapter 21) for details.] In this case, $\Sigma_t y_{it}$ is a minimal sufficient statistic for α_i, and estimation in terms of the conditional density provides a consistent estimator of β. Three other commonly used models that have this property are the Poisson and negative binomial regressions for count data [see Hausman, Hall, and Griliches (1984),[2] Cameron and Trivedi (1998), Allison (2000), Lancaster (2000) and Blundell, Griffith and Windmeijer (2002)] and the exponential regression model for a continuous nonnegative variable,

$$g(y_{it}, \beta'\mathbf{x}_{it} + \alpha_i) = (1/\lambda_{it})\exp(-y_{it}/\lambda_{it}),\ \lambda_{it} = \exp(\beta'\mathbf{x}_{it} + \alpha_i),\ y_{it} \geq 0$$

[see Munkin and Trivedi (2000)]. In all these cases, the conditional log likelihood,

$$\log L_c = \sum_{i=1}^{N} \log f\left(y_{i1}, y_{i2},..., y_{i,T_i} \,\Big|\, \left(\Sigma_{t=1}^{T_i} y_{it}\right), \mathbf{x}_{i1}, \mathbf{x}_{i2},...\right)$$

is a function of β but not α, which provides a feasible estimator of the parameters that is free of the nuisance parameters.[3] In most cases of interest to practitioners, including, for examples, those based on transformations of normally distributed variables such as the probit, tobit and truncated regression models, this method will be unusable.

2.3. Two Step Estimation

Heckman and MaCurdy (1981) suggested a 'zig-zag' sort of approach to maximization of the log likelihood function, dummy variable coefficients and all. Consider the probit model. For known set of fixed effect coefficients, $\alpha = (\alpha_1,...,\alpha_N)'$, estimation of β is straightforward. The log likelihood conditioned on these values (denoted a_i), would be

$$\log L|a_1,...,a_N = \sum_{i=1}^{N}\sum_{t=1}^{T_i} \log\Phi[(2y_{it} - 1)(\beta'\mathbf{x}_{it} + a_i)$$

This can be treated as a cross section estimation problem since with known α, there is no connection between observations even within a group. With given estimate of β (denoted **b**) the conditional log likelihood function for each α_i,

$$\log L_i|\mathbf{b} = \sum_{t=1}^{T_i} \log\Phi\left[(2y_{it} - 1)(z_{it} + \alpha_i)\right]$$

---

[2] But, see Allison (2000) for documentation of an ambiguity in the Hausman et al. formulation of the negative binomial model.

[3] Lancaster (2000) lists several cases in which the parameters of the model can be "orthogonalized," that is, transformed to a form $\alpha_i^*(\alpha,\beta)$ and β such that the log likelihood reparameterized in terms of these parameters is separable. The concentrated likelihood for the Poisson is an easily derived example. As he notes, there is no general result which produces the orthogonalization, and the number of cases is fairly small.

where $z_{it} = \mathbf{b}'\mathbf{x}_{it}$ is now a known function. Maximizing this function for each i is straightforward. Heckman and MaCurdy suggested iterating back and forth between these two estimators until convergence is achieved.[4]

There is no guarantee that this back and forth procedure will converge to the maximum of the log likelihood function because the Hessian is not block diagonal. Whether either estimator is even consistent in the dimension of N (that is, of β) even if T is large, depends on the initial estimator being consistent, and it is unclear how one should obtain that consistent initial estimator.[5] For the binary choice setting, in any group in which the dependent variable is all ones or all zeros, there is no maximum likelihood estimator for α_i - the likelihood equation for log L_i has no solution if there is no within group variation in y_{it}. This feature of the model carries over to the tobit and binomial logit models, as the authors noted and to Chamberlain's conditional logit model and the Hausman et al. estimator of the Poisson model.[6] In the Poisson and negative binomial models cases, any group which has $y_{it} = 0$ for all t contributes a zero to the log likelihood function so its group specific effect is not identified. Third, irrespective of its probability limit, the estimated standard errors for the estimator of β will be too small, again because the Hessian is not block diagonal.[7] The estimator at the β step does not obtain the correct submatrix of the information matrix.

---

[4] Polachek and Yoon (1994, 1996) applied this approach to the stochastic frontier model. See, as well, Hall (1978), Borjas and Sueyoshi (1993), Berry, Pakes and Levinsohn (1995), Petrin and Train (2002) and Greene (2002, 2003).

[5] Polachek and Yoon's (1996) application to a stochastic frontier model is based on an initial consistent estimator, OLS, so in their case, the consistency issue must be treated differently. In fact, however, though their initial estimator is consistent, subsequent iterates are not, since they are functions of the estimated fixed effects.

[6] This is not an issue in all cases, however. For example, in the linear regression model, within group variation in the dependent variable is not required for estimation of the individual constant term. In the Poisson model, estimation of α_i requires only that at least one y_{it} differ from zero.

[7] Polachek and Yoon (1996, footnote 9) argue that since the off diagonal blocks, $\mathbf{h}_{\gamma i}$ is small (T_i terms) compared to the diagonal block, this effect may be minimal. But, the missing offset to the matrix being inverted is $\Sigma_i\, \mathbf{h}_{\gamma i}$, which is of the same order.

2.4. Full Maximum Likelihood Estimation

Maximization of the log likelihood function can, in fact, be done by 'brute force,' even in the presence of possibly thousands of nuisance parameters. The strategy, which uses some well known results from matrix algebra is described in Prentice and Gloeckler (1978) [who attribute it to Rao (1973)], Chamberlain (1980, p. 227), Sueyoshi (1993) and Greene (2003). No generality is gained by treating θ separately from β, so at this point, we will simply collect them in the single $K\times 1$ parameter vector $\gamma = [\beta',\theta']'$. Denote the gradient and Hessian of the log likelihood by

$$\mathbf{g}_\gamma = \frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial \log g(y_{it},\gamma,\mathbf{x}_{it},\alpha_i)}{\partial \gamma}$$

$$g_{\alpha i} = \frac{\partial \log L}{\partial \alpha_i} = \sum_{t=1}^{T_i} \frac{\partial \log g(y_{it},\gamma,\mathrm{x}_{it},\alpha_i)}{\partial \alpha_i}$$

$$\mathbf{g}_\alpha = [g_{\alpha 1}, \ldots, g_{\alpha N}]'$$

$$\mathbf{g} = [\mathbf{g}_\gamma', \mathbf{g}_\alpha']'$$

and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\gamma\gamma} & \mathbf{h}_{\gamma 1} & \mathbf{h}_{\gamma 2} & \cdots & \mathbf{h}_{\gamma N} \\ \mathbf{h}_{\gamma 1}' & h_{11} & 0 & \cdots & 0 \\ \mathbf{h}_{\gamma 2}' & 0 & h_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \mathbf{h}_{\gamma N}' & 0 & 0 & 0 & h_{NN} \end{bmatrix}$$

where

$$\mathbf{H}_{\gamma\gamma} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial^2 \log g(y_{it},\gamma,\mathbf{x}_{it},\alpha_i)}{\partial \gamma \partial \gamma'}$$

$$\mathbf{h}_{\gamma i} = \sum_{t=1}^{T_i} \frac{\partial^2 \log g(y_{it},\gamma,\mathbf{x}_{it},\alpha_i)}{\partial \gamma \partial \alpha_i}$$

$$h_{ii} = \sum_{t=1}^{T_i} \frac{\partial^2 \log g(y_{it},\gamma,\mathbf{x}_{it},\alpha_i)}{\partial \alpha_i^2}.$$

Newton's method for computation of the parameters will use the iteration

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\alpha} \end{pmatrix}_k = \begin{pmatrix} \hat{\gamma} \\ \hat{\alpha} \end{pmatrix}_{k-1} - \mathbf{H}_{k-1}^{-1}\mathbf{g}_{k-1} = \begin{pmatrix} \hat{\gamma} \\ \hat{\alpha} \end{pmatrix}_{k-1} + \begin{pmatrix} \Delta_\gamma \\ \Delta_\alpha \end{pmatrix}.$$

By taking advantage of the sparse nature of the Hessian, this can be reduced to a computation that involves only $K\times 1$ vectors and $K\times K$ matrices;

$$\Delta_\gamma = -\left[\mathbf{H}_{\gamma\gamma} - \sum_{i=1}^{N}\left(\frac{1}{h_{ii}}\right)\mathbf{h}_{\gamma i}\mathbf{h}'_{\gamma i}\right]_{k-1}^{-1}\left(\mathbf{g}_\gamma - \sum_{i=1}^{N}\frac{g_{\alpha i}}{h_{ii}}\mathbf{h}_{\gamma i}\right)_{k-1}$$

$$= -\mathbf{H}^{\gamma\gamma}\left(\mathbf{g}_\gamma - \mathbf{H}_{\gamma\alpha}\mathbf{H}_{\alpha\alpha}^{-1}\mathbf{g}_\alpha\right)$$

$$\Delta_{\alpha i} = -\frac{1}{h_{ii}}\left(g_{\alpha i} + \mathbf{h}'_{\gamma i}\Delta_\gamma\right).$$

For a single index model, $g(y_{it}, \beta'\mathbf{x}_{it} + \alpha_i)$, with no ancillary parameters, such as the probit, logit, Poisson or exponential model, this can be written in the convenient form

$$\Delta_\gamma = \left\{\sum_{i=1}^{N}\left[\sum_{t=1}^{T_i}\psi_{it}\left(\mathbf{x}_{it}-\overline{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\overline{\mathbf{x}}_i\right)\right]\right\}^{-1} \times \left\{\sum_{i=1}^{N}\left[\sum_{t=1}^{T_i}\delta_{it}\left(\mathbf{x}_{it}-\overline{\mathbf{x}}_i\right)\right]\right\}$$

and

$$\Delta_{\alpha i} = \left[\sum_{t=1}^{T_i} -\delta_{it}/\psi_{i\cdot}\right] + \overline{\mathbf{x}}_i'\Delta_\gamma$$

where

$$\delta_{it} = \partial \log g(y_{it}, \beta'\mathbf{x}_{it} + \alpha_i)/\partial\alpha_i$$

$$\psi_{it} = \partial^2 \log g(y_{it}, \beta'\mathbf{x}_{it} + \alpha_i)/\partial\alpha_i^2$$

$$\psi_{i\cdot} = \sum_{t=1}^{T_i}\psi_{it}$$

$$\overline{\mathbf{x}}_i = \mathbf{h}_{\gamma i}/h_{ii} = \sum_{t=1}^{T_i}\psi_{it}\mathbf{x}_{it} \Big/ \sum_{t=1}^{T_i}\psi_{it}.$$

The estimator of the asymptotic covariance matrix for the slope parameters in the MLE is

$$\text{Est.Asy.Var}[\hat{\gamma}_{MLE}] = -\left[\mathbf{H}_{\gamma\gamma} - \sum_{i=1}^{N}\left(\frac{1}{h_{ii}}\right)\mathbf{h}_{\gamma i}\mathbf{h}'_{\gamma i}\right]^{-1} = -\mathbf{H}^{\gamma\gamma}$$

For the separate constant terms,

$$\text{Est.Asy.Cov}\left[a_i, a_j\right] = -\mathbf{1}(i=j)\frac{1}{h_{ii}} - \frac{1}{h_{ii}}\frac{1}{h_{jj}}\mathbf{h}'_{\gamma i}\left[\mathbf{H}_{\gamma\gamma'} - \sum_{i=1}^{N}\frac{1}{h_{ii}}\mathbf{h}_{\gamma i}\mathbf{h}'_{\gamma i}\right]^{-1}\mathbf{h}_{\gamma j}$$

$$= \frac{-\mathbf{1}(i=j)}{h_{ii}} - \left(\frac{\mathbf{h}'_{\gamma i}}{h_{ii}}\right)\mathbf{H}^{\gamma\gamma}\left(\frac{\mathbf{h}_{\gamma j}}{h_{jj}}\right).$$

For the single index model, this is

$$\text{Est.Asy.Cov}\left[a_i, a_j\right] = \frac{-\mathbf{1}(i=j)}{\psi_i} + \overline{\mathbf{x}}_i'\mathbf{V}\overline{\mathbf{x}}_j.$$

Finally,

$$\text{Est.Asy.Cov}[\hat{\gamma}_{MLE}, a_i] = \text{Est.}Asy.Var[\hat{\gamma}_{MLE}] \times \left(\frac{\mathbf{h}_{\gamma i}}{h_{ii}} \right) = -\mathbf{V}\,\overline{\mathbf{x}}_i\,.$$

Each of these involves a moderate amount of computation, but can easily be obtained with existing software and computations that are linear in N and K. Neither update vector requires storage or inversion of a $(K+N)\times(K+N)$ matrix; each is a function of sums of scalars and $K\times1$ vectors of first derivatives and mixed second derivatives. Storage requirements for α and Δ_α are linear in N, not quadratic. Even for panels of tens of thousands of units, this is well within the capacity of the current vintage of even modest desktop computers.[8] The application below, computed on an ordinary desktop computer, involves computation of a tobit model with $N = 3{,}000$.

3. Sampling Properties of the Fixed Effects Estimator

If β and θ were known, then, the MLE for α_i would be based on only the T_i observations for group i. This implies that the asymptotic variance for a_i is $O[1/T_i]$ and, since T_i is fixed, a_i is inconsistent. The estimator of β will be a function of the estimator of α_i, $a_{i,ML}$. Therefore $\mathbf{b}_{ML}$, the MLE of β is a function of a random variable which does not converge to a constant as $N \to \infty$, so neither does $\mathbf{b}_{ML}$. There may be a small sample bias as well. Andersen (1973) and Hsiao (1996) showed analytically that in a binary logit model with a single dummy variable regressor and a panel in which $T_i = 2$ for all groups, the small sample bias is +100%. Abrevaya (1997) shows that Hsiao's result extends to more general binomial logit models as long as T_i continues to equal two. Our Monte Carlo results below are consistent with this result. No general results exist for the small sample bias if T exceeds 2 or for other models. Generally accepted results are based on Heckman's (1981b) Monte Carlo study of the probit model with $T_i = 8$ and $N = 100$ in which the bias of the

[8] Sueyoshi (1993) after deriving these results expressed some surprise that they had not been incorporated in commercial software. As of this writing, it appears that LIMDEP [Econometric Software, (2003)] is still the only package that has done so.

slope estimator was toward zero (in contrast to Hsiao) and on the order of only 10%. On this basis, it is often suggested that in samples at least this large, the small sample bias is probably not too severe. However, our results below suggest that the pattern of overestimation in the probit model persists to larger T as well, and Heckman's results appear to be too optimistic. Neyman and Scott (1948) are often invoked to assert the extension of this result to other models as well. In point of fact, Neyman and Scott did not claim any generality for the small sample bias of the maximum likelihood estimator; they observed it in passing in one narrow case (the variance of the fixed effects estimator in a model with no regressors) during the course of their examination of the asymptotic efficiency of the MLE in the presence of the nuisance parameters. As we find below, there appears to be no predictable pattern to the sign, or even the presence of a small sample bias of the fixed effects estimator.

3.1. Discrete Choice Models

The experimental design for Heckman's Monte Carlo analysis of the fixed effects probit estimator was as follows:

$$Y_{it} = \sigma_\tau \tau_i + \beta z_{it} + \varepsilon_{it},\ i = 1,\ldots,100,\ t = 1,\ldots,8,$$

$$\tau_i \sim N[0,1],$$

$$z_{it} = 0.1t + 0.5z_{i,t-1} + U_{it},\ U_{it} \sim U[-0.5,0.5],\ z_{i0} = 5 + 10.0U_{i0},$$

$$\varepsilon_{it} \sim N[0,1],$$

$$y_{it} = \mathbf{1}[Y_{it} > 0].$$

[The initialization of z_{it} is given in Nerlove (1971).] Heckman's results are summarized in Table 1. For the case of interest here, his results for the probit model with $N = 100$ and $T = 8$ suggest, in contrast to the evidence for the logit model, a slight *downward* bias in the slope estimator. The striking feature of his results is how small the bias seems to be even with T as small as 8.

We have been unable to replicate any of Heckman's results. Both his and our own results with his experimental design are shown in Table 1. Some of the difference can be explained by different random number generators. But, this would only explain a small part of the strikingly different outcomes of the experiments and not the direction. In contrast to Heckman, using his specification, we find that the probit estimator, like the logit estimator, is substantially biased *away* from zero when $T = 8$. Consistent with expectations, the bias is far less than the 100% that appears when $T = 2$. The table contains three sets of results. The first are Heckman's reported values. The second and third sets of results are our computations for the same study. Heckman based his conclusions on 25 replications. To control the possibility that some of the variation is due to small sample effects, we have redone the analysis using 100 replications. The results in the second and third row of each cell are strongly consistent with the familiar results for the logit model and with our additional results discussed below. The bias in the fixed effects estimator appears to be quite large, and, in contrast to Heckman's results, is away from zero in all cases. The proportional bias does not appear to be a function of the parameter value.

Table 1. Heckman's Monte Carlo Study of the Fixed Effects Probit Estimator

	β = 1.0	**β = -0.1**	**β = -1.0**
$\sigma_\tau^2 = 3$	0.90[a] 1.286[b] 1.240[c]	-0.10 -0.1314 -0.1100	-0.94 -1.247 -1.224
$\sigma_\tau^2 = 1$	0.91 1.285 1.242	-0.09 -0.1157 -0.1127	-0.95 -1.198 -1.200
$\sigma_\tau^2 = 0.5$	0.93 1.213 1.225	-0.10 -0.1138 -0.1230	-0.96 -1.199 -1.185

[a]Reported in Heckman (1981), page 191.
[b]Mean of 25 replications
[c]Mean of 100 replications

There is an important shortcoming in the design of the foregoing experiment. The underlying model is not a fixed effects model; it is a random effects model. The signature feature of the fixed effects model is correlation between the effects and the included variables, and by construction, there is none between τ_i and z_{it} in the model above. As such, the foregoing does not give evidence on the point for which it is usually cited, that is, the small sample bias of the unconditional fixed effects estimator *of the fixed effects model*. More to the point, if the researcher knows that the effects are not correlated with the included variables, then a random effects approach should be preferable, and the issue at hand becomes whether the normal distribution typically assumed is a valid assumption and what are the implications if it is not. Current technology provides a variety of useful approaches for random effects and random parameters models when it can be assumed that the effects and the included variables are orthogonal.

We will examine the behavior of the estimator in somewhat greater detail. We are interested in whether Hsiao's result carries over to other models, and how Heckman's results change when T is not equal to 8. We will examine several index function models, the binomial logit, binomial probit, ordered probit, tobit, truncated regression and Weibull models. (The continuous choice models are considered in the next section.) The experiment is designed as follows: All models are based on the same index function:

$$w_{it} = \alpha_i + \beta x_{it} + \delta d_{it},$$

where $\beta = \delta = 1$,

$$x_{it} \sim N[0,1^2]$$

$$d_{it} = \mathbf{1}[x_{it} + h_{it} > 0] \text{ where } h_{it} \sim N[0,1^2]$$

$$\alpha_i = \sqrt{T}\bar{x}_i + a_i,\ a_i \sim N[0,1^2]$$

In all cases, we estimate the two coefficients on x_{it} and d_{it}, where both coefficients equal 1.0, and the fixed effects (which are not used or presented below). The correlations between the variables are approximately 0.7 between x_{it} and d_{it} , 0.4 between α_i and x_{it} and 0.2 between α_i and d_{it}. The data generating processes examined here are as follows:

Probit: $y_{it} = \mathbf{1}[w_{it} + \varepsilon_{it} > 0]$,

Ordered Probit: $y_{it} = \mathbf{1}[w_{it} + \varepsilon_{it} > 0] + \mathbf{1}[w_{it} + \varepsilon_{it} > 3]$,

Logit: $y_{it} = \mathbf{1}[w_{it} + v_{it} > 0]$, $v_{it} = \log[u_{it}/(1-u_{it})]$,

where $\varepsilon_{it} \sim N[0,1^2]$ denotes a draw from the standard normal population and $u_{it} \sim U[0,1]$ denotes a draw from the standard uniform population. Models were fit with $T = (2, 3, 5, 8, 10, 20)$ and with

N = (100, 500, 1,000). (Note that this includes Heckman's experiment.) Each model specification, group size, and number of groups was fit 200 times with random draws for ε_{it} or u_{it}. For purposes of our analysis, we based conclusions on the N = 1,000 experiments. The conditioning data, x_{it}, d_{it} and α_i are held constant. The full set of parameters, including the dummy variable coefficients, are estimated using the results given earlier. For each of the specifications listed, properties of the sampling distribution are estimated using the 200 observations on β and δ.[9]

Table 2 lists the means of the empirical sampling distribution for the three different discrete choice estimators for the samples of 1,000 individuals. At this point, we are only interested in the mean of the sampling distribution as a function of T, so we use only the results based on the largest (N) samples. The bias of the fixed effects estimator in the binary and ordered choice models is large and persistent. Even at T = 20, we find substantial biases. With T = 2, the Anderson/Hsiao result is clearly evident, even more so in the ordered probit model. Increasing the sample size (N) from 100 to 1,000 did nothing to remove this effect, but the increase in group size (T) from 2 to 20 has a very large effect. We conclude that this is a persistent bias that can, indeed, be attributed to the "small T problem." The results for the probit model with T = 8 are the counterparts to Heckman's results. The biases in Table 2 are quite unlike those in his study. The ordered probit model, which has not been examined previously, shows the same characteristic pattern as the binomial models.

Table 2. Means of Empirical Sampling Distributions, *N* = 1000 Individuals Based on 200 Replications.

	***T*=2**	***T*=3**	***T*=5**	***T*=8**	***T*=10**	***T*=20**
	β δ	β δ	β δ	β δ	β δ	β δ
Logit Coeff	2.020, 2.027	1.698, 1.668	1.379, 1.323	1.217, 1.156	1.161, 1.135	1.069, 1.062
Logit M.E.	1.676 1.660	1.523 1.477	1.319 1.254	1.191 1.128	1.140 1.111	1.034 1.052
Probit Coeff	2.083, 1.938	1.821, 1.777	1.589, 1.407	1.328, 1.243	1.247, 1.169	1.108, 1.068
Probit M.E.	1.474 1.388	1.392 1.354	1.406 1.231	1.241 1.152	1.190 1.110	1.088 1.047
Ord. Probit	2.328, 2.605	1.592, 1.806	1.305, 1.415	1.166, 1.220	1.131, 1.158	1.058, 1.068

[9] A similar study over a range of group sizes is carried out for the binary logit model by Katz (2001).

The focus on coefficient estimation in these models overlooks an important aspect of estimation in a binary choice model. Unless one is only interested in signs and statistical significance (and, if so, then the incidental parameters problem may be a moot point), then the relevant object of estimation in the model is the marginal effect, not the coefficient itself. For the two binary choice models, the marginal effects are

$$\frac{\partial E[y_{it} \mid \alpha_i, x_{it}, d_{it}]}{\partial x_{it}} = \beta f(\alpha_i + \beta x_{it} + \delta d_{it})$$

for the continuous variable x_{it} and

$$\Delta E[y_{it}|\alpha_i, x_{it}, d_i] = F(\alpha_i + \beta x_{it} + \delta) - F(\alpha_i + \beta x_{it})$$

for the dummy variable d_{it}, where $f(.)$ and $F(.)$ denote the density and CDF (normal or logistic), respectively. These are functions of the data, so there is in principle no 'true' value to be estimated. But, these are typically computed at the means of the independent variables. Taking this as our benchmark, the estimated values would be based on averages of zero for α_i and x_{it} and 0.5 for d_{it}. The 'true' marginal effects would be $1\times\phi(0 + 1\times0 + 1\times.5) = 0.352$ and $\Phi(1) - \Phi(0) = 0.3413$ for the probit model and $1\times\Lambda(.5)[1-\Lambda(.5)] = 0.235$ and $\Lambda(1) - \Lambda(0) = .231$ for the logit model for x_{it} and d_{it}, respectively. The estimated values would be obtained by inserting the estimated coefficients in the preceding expressions. In each case, the overestimated coefficient acts to increase the multiplier but attenuate the scale factor, so the relationship between the marginal effects and the coefficients is unclear. The second row of values for the logit and probit models in Table 2 gives the ratio of what would be the estimated marginal effect to the 'true' marginal effects for the logit and probit models. Comparison of the entries suggests that the biases are comparable for $T \geq 5$. However, the first two columns suggest that the commonly accepted result of a 100% bias when $T = 2$ substantially overstates the case. The bias is still large, but well under 100%. In all cases save for the last, the marginal effect is closer to the true value than the coefficient estimator is to its population counterpart. We do note, these results do not redeem the estimator. However, they do cast some new light on a long held result, the bias for $T = 2$.

The preceding analysis and its counterpart elsewhere in the literature leaves an open question. Believing that the fixed effects model is appropriate for their data, but faced with the foregoing results, the analyst committed to a parametric approach has (at least) three alternatives: use the fixed effects estimator in spite of the incidental parameters issue, use the random effects estimator, even though it is, at least in principle, inconsistent, or ignore the heterogeneity and use the pooled estimator. It is unclear which should be preferred. All three estimators are biased and inconsistent. Table 3 presents a comparison of these three estimators for the same sample design for the probit model with $T = 3$ and $T = 8$, with $N = 1{,}000$. All three estimators were computed with the same data, 200 times. The table lists the sample means and the root mean squared deviations around the true values of 1.0 for β and δ. For which among the three to choose, it is clear that the random effects estimator is overwhelmingly the worst of the three. It is ambiguous whether one should use the fixed effects estimator or pool the data and ignore the heterogeneity. The interesting result is that while the fixed effects estimator is biased upward, the pooled estimator is biased downward. For the worse case, $T = 3$, the bias of the pooled estimator is considerably smaller and the root mean squared deviation is as well. For $T = 3$, without question, the pooled estimator is superior. For $T = 8$, it is unclear. In this case, the biases are opposite, but comparable. The root mean squared error for β favors the fixed effects estimator while that for δ favors the pooled estimator. The comparison is unclear. It seems likely based on this and all the preceding results that for T larger than 8, the results will probably favor the fixed effects estimator. On the other hand, it is obvious that the better course when T is very small (between the two problematic ones) is the pooled estimator.

Table 3. Means and Root Mean Squared Deviations of Fixed Effects, Random Effects and Pooled Estimators for the Probit Model

	$T = 3$				$T = 8$			
	β		δ		β		δ	
	Mean	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
Pooled	0.953	0.671	0.655	0.349	0.797	0.204	0.604	0.397
Random	0.415	0.588	2.629	1.634	0.249	0.752	2.286	1.288
Fixed	1.868	0.909	1.769	0.839	1.332	0.340	1.236	0.262

3.2. The Tobit, Truncated Regression and Weibull Models

The tobit model was simulated using the same experimental design, with replication

$$y_{it} = \mathbf{1}[c_{it} > 0] \times c_{it},\ c_{it} = w_{it} + \varepsilon_{it}.$$

Table 4 presents the simulation results for the tobit model specified above. It appears that the tobit fixed effects estimator is not biased at all. The result is all the more noteworthy in that in each data set, roughly 40 - 50% of the observations are censored. If none of the observations were censored, this would be a linear regression model, and the resulting OLS estimator would be the consistent linear LSDV estimator. But, with roughly 40% of the observations censored, this is a quite unexpected result. However, the average of the 200 estimates of σ - the true value is also 1.0 - shows that the incidental parameters problem shows up in a different place here. The estimated standard deviation is biased *downward*, though with a bias that does diminish substantially as T increases. This result is not innocuous. Consider estimating the marginal effects in the tobit model with these results. In general in the tobit model, for a continuous variable, $\delta_k = \partial E[y_i|\mathbf{x}_i]/\partial x_{ik} = \beta_k \times \Phi(\beta'\mathbf{x}_i/\sigma)$ where $\Phi(z)$ is the cdf of the standard normal distribution. This is frequently computed at the sample means of the data. Based on our experimental design, the overall means of the variables would be zero for α_i and x_i and 0.5 for d_i. Therefore, the scale factor estimated, using the true values of the slope parameters as they are (apparently) estimated consistently, would be $\Phi(0.5/\hat{\sigma})$. The ratio of this value computed at the average estimate of σ to the value computed at σ = 1 [which would be Φ(.5) = .6914] is given in the last row of the table, where it can be seen that for small T, there is some *upward* bias in the marginal effects, but far less than that in the discrete choice models. On the other hand, at $T = 8$ (Heckman's case), the tobit model appears to be essentially consistently estimated in spite of the incidental parameters issue. It is tempting to invoke Neyman and Scott's result mentioned earlier to explain this finding, but the censoring aspect of the model and the contradictory results below for the truncation model suggest that would be inappropriate.

Table 4. Means of Empirical Sampling Distributions, Tobit, Truncated Regression and Weibull Models, *N* = 1000 Individuals Based on 200 Replications.

	T=2	T=3	T=5	T=8	T=10	T=20
			Tobit Model			
β	0.991	0.985	0.997	1.000	1.001	1.008
δ	1.083	0.991	1.010	1.008	1.004	1.00
σ	0.644	0.768	0.864	0.914	0.928	0.964
Scale factor	1.13	1.07	1.04	1.02	1.01	1.02
			Truncated Regression Model			
β	0.892	0.921	0.955	0.967	0.971	0.986
δ	0.740	0.839	0.888	0.934	0.944	0.973
σ	0.664	0.782	0.869	0.920	0.935	0.968
Scale factor	1.033	1.021	1.006	1.004	1.0003	1.001
Mar.Effect	0.448	0.457	0.467	0.472	0.474	0.480
			Weibull Duration Model			
β	0.706	0.773	0.806	0.832	0.836	0.861
δ	1.284	1.207	1.170	1.128	1.117	1.085
σ	0.512	0.659	0.767	0.826	0.847	0.878

The truncated regression model is generated by the nonlimit observations in the censored regression setting. [See Hausman and Wise (1977).] Thus, for the simple case of lower truncation at zero (any other point, or upper truncation is a trivial modification of the model),

$$y_{it}^* = \alpha_i + \beta x_{it} + \delta d_{it}, + \varepsilon_{it}$$

$$y_{it} = y_{it}^* \text{ if } y_{it}^* > 0 \text{ and is unobserved otherwise.}$$

The log likelihood for the truncated regression model is

$$\log L = \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \log \left[\frac{1}{\sigma} \phi \left(\frac{y_{it} - \alpha_i - \beta x_{it} - \delta d_{it}}{\sigma} \right) \right] - \log \Phi \left[\frac{\alpha_i + \beta x_{it} + \delta d_{it}}{\sigma} \right] \right\}.$$

Based on results already obtained, we can deduce how the MLE in this model is likely to behave. By adding and subtracting a term and using the symmetry of the normal distribution, the log likelihood function for the tobit model may be written

$$\begin{aligned}\log L &= \sum\nolimits_{i,t,y>0} \log\left[\frac{1}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\right] + \sum\nolimits_{i,t,y=0} \log\Phi\left(\frac{-\beta'\mathbf{x}_{it}}{\sigma}\right) \\ &= \left\{\sum\nolimits_{i,t,y>0} \log\left[\frac{1}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\right] - \sum\nolimits_{i,t,y>0} \log\Phi\left(\frac{\beta'\mathbf{x}_{it}}{\sigma}\right)\right\} \\ &+ \left\{\sum\nolimits_{i,t,y=0} \log\Phi\left(\frac{-\beta'\mathbf{x}_{it}}{\sigma}\right) + \sum\nolimits_{i,t,y>0} \log\Phi\left(\frac{\beta'\mathbf{x}_{it}}{\sigma}\right)\right\}\end{aligned}$$

The first line of the result is the log likelihood for a truncated regression model for the nonlimit observations. The second line is the log likelihood for the binary probit model. Since $\sigma = 1$ (though the more general case produces the same result), we can see that since the tobit estimator of the slopes is unbiased, and the probit estimator is biased upward, we should expect the truncated regression estimator to be biased downward, toward zero. The results in Table 4 are consistent with this observation.

The simulations for the truncated regression model are produced using Geweke's (1986) suggested method,

$$y_{it} = \alpha_i + \beta x_{it} + \delta d_{it}, + \sigma\Phi^{-1}\{u_{it} + (1-u_{it})\Phi[(\alpha_i + \beta x_{it} + \delta d_{it})/\sigma]\}$$

where u_{it} is a draw from the standard uniform population. This one to one transformation produces a single draw from the truncated at zero normal distribution with mean $\alpha_i + \beta x_{it} + \delta d_{it}$, and standard σ. The conditional mean function in the truncated regression model is

$$\begin{aligned}E[y_{it}|\alpha_i, x_{it}, d_{it}] &= \alpha_i + \beta x_{it} + \delta d_{it}, + \sigma\lambda[(\alpha_i + \beta x_{it} + \delta d_{it})/\sigma] \\ &= \alpha_i + \beta x_{it} + \delta d_{it}, + \sigma\lambda_{it}\end{aligned}$$

where $\lambda(z) = \phi(z)/\Phi(z)$. For a continuous variable, x_{it}

$$\frac{\partial E[y_{it} \mid \alpha_i, x_{it}, d_{it}]}{\partial x_{it}} = \beta\left[1 - \lambda_{it}\left(\frac{\alpha_i + \beta x_{it} + \delta d_{it}}{\sigma} + \lambda_{it}\right)\right],$$

so, for estimating partial effects, the necessary scale factor is the term in square brackets. [The term is bounded by zero and one. See, e.g., Maddala (1983) or Greene (2003, Section 22.2.3).] Once again, the 'true' value would depend on the data. Repeating the logic used for the tobit model, we evaluated this at the true values of $\alpha_i = x_{it} = 0$ and $\delta d_{it} = 1(.5)$ with $\sigma = 1$, so that our population value is 0.4862. The sample estimates would be based on $\hat{\delta}(.5)/\hat{\sigma}$. As before, the scale factor in the table displays the average scale factor divided by the true value as well as the estimated marginal effect, now the scale factor times the estimated coefficient. Though the coefficients and the estimated standard deviation in this model are noticeably biased, the effects are largely offsetting in the marginal effects, which are quite close to the true value for all sample sizes.

Several panel data duration models have been analyzed in this setting as well. Chamberlain (1985) analyzed the Weibull and gamma models and showed how the fixed effects could be conditioned out of the models by analyzing $\log(y_{it}/y_{i1})$. Using Kalbfleisch and Prentice's (1980) formulation of the model, we have the survival function

$$S(y_{it}|\alpha_i,x_{it},d_{it}) = \exp[-(\lambda_{it}y_{it})^p],\ \lambda_{it} = \exp[-(\alpha_i + \beta x_{it} + \delta d_{it})],\ p = 1/\sigma$$

and hazard function

$$h(y_{it}|\alpha_i,x_{it},d_{it}) = \lambda_{it}p(\lambda_{it}y_{it})^{p-1}.$$ [10]

Duration data are often censored. Let $Q_{it} = 1$ if the observation is 'complete' and $Q_{it} = 0$ if the observation is censored. Then, the log likelihood function is

$$\log L = \sum_{i,t}\left[\log S(y_{it}\,|\,\alpha_i,x_{it},d_{it}) + Q_{it}\log h(y_{it}\,|\,\alpha_i,x_{it},d_{it})\right]$$

Replications for the simulations are drawn by inverting the survival function to produce draws

$$\log y_{it} = \alpha_i + \beta x_{it} + \delta d_{it} + \sigma\log(-\log(1-u_{it})).$$

[10] This form reparameterizes both Chamberlain's and Lancaster's description of the model. In the former, Chamberlain has dropped the log of the scale parameter from the log of the hazard, but nothing is lost if it is simply absorbed into the fixed effect.

Observations on $\log y_{it}$ were censored at 3. Once again, all three structural parameters of the model are equal to 1.0. Table 4 presents the estimates for the Weibull model with censored data. In this instance, the two estimators of β and δ are converging to their population values from different directions, β from below and δ from above. As in the tobit case, the estimator of σ is attenuated.[11] These results for the slopes are actually contradictory if we view the Weibull model with censoring as a distributional alternative to the tobit model. Evidently, the structure is more complicated than that.

These findings highlight two results. First, it is clear that the results for the binary choice models do not carry over to these continuous choice models. Indeed, there is no persistent pattern whether the estimator is biased upward or downward, or at all in these settings. Where there is a finite sample bias, it appears to be much smaller than for the probit and logit estimators. Second, they suggest the ambiguity of focusing on the slope coefficients in estimation of these models. One might be tempted to conclude that the fixed effects estimator is unbiased in the tobit setting - by dint of only the coefficients, it appears to be. But, when the slopes of the model are computed, the force of the small sample bias is exerted on the results through the disturbance standard deviation. Third, however, the results in Table 4 suggest that the conventional wisdom on the fixed effects estimator, which has been driven by the binary choice models, might be too pessimistic. With T equal to only 5, the estimators appears to be only slightly affected by the incidental parameters problem. Even at $T = 3$, the 7% upward bias in the marginal effects in the tobit model is likely to be well within the range of the sampling variability of the estimated parameter.

[11] Lancaster (2000, p. 397) states "the estimate for θ converging to a number less than the true value." In his formulation, θ is $1/\sigma$ for the formulation above, so our results are not consistent with his assertion. The text seems to suggest Chamberlain as the source of the claim, but Chamberlain does not discuss the issue, so this inconsistency is unresolved.

3.3. Estimated Standard Errors

In all the cases examined, a central issue is the extra variation induced in the parameter estimators by the presence of the inconsistent fixed effect estimators. Since the estimator, itself, is inconsistent, one should expect distortions in estimators of the asymptotic covariance matrix. Table 5 lists, for each model, the estimated asymptotic standard errors computed using the estimated second derivatives matrix and the empirical standard deviation based on the 200 replications in the simulation, using the $N = 1000$, $T = 8$ group of estimators. The 'analytic' estimator is obtained by averaging the 200 estimated asymptotic standard errors. The empirical estimator is the sample standard deviation of the 200 estimates obtained in the simulation. The latter should give a more accurate assessment of the sampling variation of the estimator while the former is, itself, an estimator which is affected by the incidental parameters problem. There is clearly some downward bias in almost all the estimated standard errors. The implication is that as a general result, test statistics such as the Wald statistics (t ratios) will tend to be too large when based on the analytic estimator of the asymptotic variance – estimates are biased upward and standard errors are biased downward. The last two columns in the table give the percentage by which the diagonals of the inverse of the Hessian underestimate the sampling variance of the estimator.

Table 5. Estimated Standard Errors and Sample Standard Deviations of Sample Estimates

	Analytic		Empirical		% Underestimate	
Model	β	δ	β	δ	β	δ
Probit	0.2234	0.3008	0.2606	0.3254	14.0	7.6
Logit	0.2324	0.3697	0.2627	0.4312	11.5	14.3
Ordered Probit	0.1281	0.2088	0.1487	0.2392	13.9	12.7
Tobit	0.0692	0.1296	0.0800	0.1386	13.5	6.5
Truncation	0.0242	0.0476	0.0265	0.0431	8.7	-10.4
Weibull	0.0175	0.0350	0.0181	0.0375	3.3	6.7

3.4. Consistency and Relative Inefficiency in the Tobit Model

We consider, finally, the relative (in)efficiency of the estimator in a model in which there appears to be little or no bias in the coefficient estimator, the tobit model. The model is respecified as

$$y_{it}^* = \alpha + \beta x_{it} + \delta d_{it} + \varepsilon_{it}, \alpha = 0, \beta = 1, \delta = 1, \sigma = 1, \varepsilon_{it} \sim N[0,\sigma^2],$$

$$y_{it} = 1(y_{it}^* > 0)y_{it}^*$$

Thus, the true constant terms are all equal to each other and equal zero. In this setting, the maximum likelihood estimator should enjoy a decided advantage over the fixed effects estimator. Table 6 below lists the ratio of the empirical variances of the two estimators of β for values of N of 50, 100, 250, 500, 1,000 and 3,000 and T = 2, 3, 5, 8, 10, and 20. (The results for estimation of δ are quite similar and are omitted for brevity.) As before, each configuration is replicated 200 times using the same sample design as employed earlier. The table gives the ratio of the sample variances of the 200 observations on each estimator. (The bias in the analytic standard errors examined in the preceding section is thereby avoided.) The table contains several surprising results. First, given the number of superfluous parameters that are folded into the mix, the relative inefficiency of the fixed effects maximum likelihood estimator is surprisingly small. With T = 8 and N = 100 (Heckman's case), the estimated standard deviation of the fixed effects estimator exceeds that of the MLE by only about 10%. What is more surprising, however, is the behavior of the ratio as N increases. If the fixed effects estimator is inconsistent, while by the usual results for maximum likelihood estimation, the MLE is consistent, then as N increases, this ratio should be increasing. If the asymptotic variance of the MLE is converging to zero and the counterpart for the fixed effects estimator is not, this ratio should be diverging. With the exception of the apparently special case of T = 2, it is not - it is relatively fixed, then falling for the larger values of N. It is, therefore, unclear what is meant by "inconsistency," at least in the setting of the tobit model. The estimator is unbiased and apparently its sampling variance is declining to that of the constrained MLE and, as a consequence, is declining absolutely as well! We note a third point not pursued here. A considerable amount of research has been done to divine estimators for this model that bypass the (now ambiguous) incidental parameters problem. It would be useful to see if there is a practical payoff, that is, how the variation in a semiparametric estimator would compare to the maximum likelihood fixed effects estimator. (The robustness of the semiparametric formulations to, e.g., nonnormality of the underlying distribution or to heteroscedasticity is duly noted. Indeed, it might be useful to measure the behavior of the 'robust' estimator against the maximum likelihood estimator in the presence of the failures of the narrow assumptions.)

Table 6. Ratio of Sample Variances of Fixed Effects and Maximum Likelihood of Estimates of β

	***T*=2**	***T*=3**	***T*=5**	***T*=8**	***T*=10**	***T*=20**
***N*=50**	1.80	1.62	1.30	1.24	1.17	1.03
***N*=100**	2.26	1.55	1.35	1.20	1.16	1.11
***N*=250**	2.27	1.84	1.40	1.22	1.15	1.05
***N*=500**	2.58	1.81	1.32	1.22	1.11	1.04
***N*=1,000**	2.67	1.39	1.15	1.14	1.09	1.02
***N*=3,000**	2.30	1.34	1.16	1.13	1.04	1.00

4. Conclusions

The Monte Carlo results obtained here suggest a number of conclusions. As widely believed, the fixed effects estimator shows a large finite sample bias in discrete choice models when T is very small. The general results for the probit and logit models are mimicked by the ordered probit model. The bias is persistent, but it does drop off rapidly as T increases to 3 and more. Heckman's widely cited result for the probit model appears to be incorrect, however. The discrepancy does not appear to be a function of the mechanism used to generate the exogenous variables. Heckman used Nerlove's (1971) dynamic model whereas we used essentially a random cross section. Results were similar for the two cases. The extreme result usually cited for the binary choice model with $T = 2$ may itself be a bit of an exaggeration. The marginal effects in these models are overestimated by a factor closer to 50%. A result which has not been considered previously is the incidental parameters effect on estimates of the standard errors of the maximum likelihood estimators. We find that while the coefficients are uniformly overestimated, the asymptotic variances are generally underestimated.

Models with mixed and continuous dependent variables behave quite differently from the discrete choice models. Overall, where there are biases in the estimates, they are much smaller than in the discrete choice models. The estimator shows essentially no bias in the slope estimators of the tobit model. But, the small sample bias appears to show up in the estimate of the disturbance variance. This bias would be transmitted to estimates of marginal effects. However, this bias appears to be small if T is 5 or more. The truncated regression and Weibull models are contradictory, and make it clear that the direction of bias in the fixed effects model is model specific. It is downward in the truncated regression and in either direction in the Weibull model.

We also considered the behavior of the fixed effects estimator measured against the consistent and fully efficient maximum likelihood estimator in the tobit model. Here the results are quite surprising. First, there is little question that the fixed effects estimator of the location coefficients in the model (β) display no significant bias. When we compare the asymptotic variances of the two estimators the results suggest first, that relatively little efficiency loss is occurring. Second, surprisingly, the relative variance of the fixed effects estimator is not increasing as one might expect; it is *decreasing* as a function of *N*. The description of the fixed effects estimator as "inconsistent of order 1/T" at the very best, is misleading in this context; at worst, it is incorrect. The results certainly seem to suggest that for this model, the fixed effects estimator does converge to its expectation. This is a puzzling and anomalous result. Finally, we submit that it remains to consider what would result from a like comparison of recently developed semiparametric estimators of the same parameters that have been proposed to avoid estimation of the fixed effects coefficients.

The received studies of the fixed effects estimator have focused intensively and exclusively on the probit and logit binary choice models. The technology exists to estimate fixed effects models in many other settings. Given the availability of high quality panel data sets, there should be substantial payoff to further scrutiny of this useful model in settings other than the binary choice models. The question does remain, should one use this technique? It obviously depends on *T* and the model in question. The reflexive negative reaction, however, because it 'biased and inconsistent' neglects a number of considerations, and might be ill advised if the alternative is a random effects approach or a semiparametric approach which sacrifices most of the interesting content of the analysis in the interest of robustness. The preceding suggests that some further research on the subject would be useful. Lancaster (2000, fn 18) notes "The fact that the inconsistency of ML in these models [Neyman and Scott's simple regression models] is rather trivial has been unfortunate since it has, I think, obscured the general pervasiveness and difficulty of the incidental parameters problem in econometric models." The results obtained here are offered in strong support of his assertion.

References

Abrevaya, J., "The Equivalence of Two Estimators of the Fixed Effects Logit Model," *Economics Letters* 55, 1, 1997, pp. 41-44.

Ahn, S. and P. Schmidt, "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics,* 68, 1995, pp. 3-38.

Allison, P., "Problems with the Fixed-Effects Negative Binomial Models," Manuscript, Department of Sociology, University of Pennsylvania, 2000.

Andersen, E., *Conditional Inference and Models for Measuring.* Mentalhygiejnisk Forsknings Institut: Copenhagen, 1973.

Arellano, M. and S. Bond. , "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58, 1991, pp. 277-297.

Arellano, M. and O. Bover, "Another Look at the Instrumental Variable Estimation of Error Components Models. *Journal of Econometrics* 68, 1995, pp.29-51.

Arellano, M. and B. Honore, "Panel Data Models: Some Recent Developments," in E. Leamer and J. Heckman, eds., *Handbook of Econometrics, Volume 5*, Amsterdam, North Holland, 2001, pp. 3229-3296.

Baltagi, B., *Econometric Analysis of Panel Data, 2nd ed.*, New York, John Wiley and Sons, 2000.

Berry, S., J. Levinsohn and A. Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 1995, pp. 841-889.

Blundell, R., R. Griffith and F. Windmeijer, "Individual Effects and Dynamics in Count Data Models," Journal of Econometrics, 108, 1, 2002, pp. 113-131.

Borjas, G. and G. Sueyoshi, "A Two-Stage Estimator for Probit Models with Structural Group Effects," *Journal of Econometrics*, 64, 1/2, 1994, pp. 165-182.

Cameron, C. and P. Trivedi, *Regression Analysis of Count Data*, New York, Cambridge University Press, 1998.

Cerro, J., "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," Manuscript, CEMFI, 2002.

Chamberlain, G., "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* 47, 1980, pp. 225-238.

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

Charlier, C., B. Melenberg and A. van Soest, "A Smoothed Maximum Score Estimator for the Binary Choice Panel Data Model and an Application to Labor Force Participation," *Statistica Neerlandica*, 49, 1995, pp. 324-342.

Chen, X., J. Heckman and E. Vytlacil, "Identification and Root-N Efficient Estimation of Semiparametric Panel Data Models with Binary Dependent Variables and a Latent Factor," Manuscript, Department of Economics, University of Chicago, 1999.

Econometric Software, Inc., *LIMDEP, Version 8.0*, Plainview, New York, Econometric Software, 2003.

Geweke, J., "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics*, 2, 1986, pp. 127-142.

Greene, W., "Fixed and Random Effects in Stochastic Frontier Models," Working Paper #02-16, Department of Economics, Stern School of Business, New York University, 2002.

Greene, W., *Econometric Analysis*, *5th ed.* Prentice Hall, Englewood Cliffs, 2003

Hall, R., "A General Framework for Time Series - Cross Section Estimation, *Annales de l'INSEE*, 30/31, 1978, pp. 177-202.

Hahn, J., "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects," *Econometric Theory*, 17, 2001, pp. 913-932.

Hausman, J., B. Hall and Z. Griliches, "Econometric Models for Count Data with an Application to the Patents - R&D Relationship," *Econometrica* 52, 1984, pp. 909-938.

Hausman, J. and D. Wise, "Social Experimentation, Truncated Distributions, and Efficient Estimation," Econometrica, 45, 1977, pp. 919-938.

Heckman, J., "Simple Statistical Models for Discrete Panel Data Developed and Applied to Tests of the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence," *Annales de l'INSEE*, 30/31, 1978, pp. 227-269.

Heckman, J., "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time - Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Econometric Applications*, Manski, C. and McFadden D. (eds.), Cambridge, MIT Press, 1981a.

Heckman, J., "Statistical Models for Discrete Panel Data. In *Structural Analysis of Discrete Data with Econometric Applications*, Manski, C. and McFadden D. (eds.), Cambridge MIT Press., 1981b.

Heckman, J. and T. MaCurdy, "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies* 47, 1981, pp.247-283.

Honore, B. and T. Kyriazidou., "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 4, 2000, pp. 839-874.

Honore, B. and A. Lewbel, "Semiparametric Binary Choice Panel Data Models without Strictly Exogenous Regressors," *Econometrica*, 70, 2002, pp. 2053-2063.

Hsiao, C. "Logit and Probit Models. In *The Econometrics of Panel Data: Handbook of Theory and Applications, Second Revised Edition,* Matyas, L. and Sevestre, P. (eds.), Dordrecht, Kluwer Academic Publishers, 1996.

Kalbfleisch, J. and R. Prentice, *The Statistical Analysis of Failure Time Data*, New York, John Wiley and Sons, 1980.

Katz, E., "Bias in Conditional and Unconditional Fixed Effects Logit Estimation," *Political Analysis*, 9, 4, 2001, pp. 379-384.

Krailo, M. and M. Pike., "Conditional Multivariate Logistic Analysis of Stratified Case Control Studies," *Applied Statistics* 44, 1984, pp.95-103.

Laisney, F. and M. Lechner, "Almost Consistent Estimation of Panel Probit Models with 'Small' Fixed Effects," University of St. Gallen, Department of Economics, Working Paper 2002-15, 2002.

Lancaster, T., "Panel Binary Choice with Fixed Effects," Manuscript, Department of Economics, Brown University, 1999.

Lancaster, T. 2000. The incidental parameters problem since 1948. *Journal of Econometrics*, 95: 391-414.

Magnac, T., "Binary Variables and Fixed Effects: Generalizing the Conditional Logit Model," manuscript, INRA and CREST, Paris, 2002.

Maddala, G., *Limited Dependent and Qualitative Variables in Economet*rics, Cambridge University Press, New York, 1983.

Maddala, G., "Limited Dependent Variable Models Using Panel Data. *Journal of Human Resources*, 22, 1987, pp. 307-338.

Manski, C., "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 1987, pp. 357-362.

Nerlove, M., "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections., *Econometrica* 39, 1971, pp. 359-382.

Neyman, J. and E. Scott., "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1948, pp. 1-32.

Munkin, M. and P. Trivedi., "Econometric Analysis of a Self Selection Model with Multiple Outcomes Using Simulation Based Estimation: An Application to the Demand for Health Care," Manuscript, Department of Economics, Indiana University, 2000.

Orme, C., Two-Step Inference in Dynamic Non-Linear Panel Data Models," Manuscript, School of Economic Studies, University of Manchester, 1999.

Petrin, A. and K. Train, "Omitted Product Attributes in Discrete Choice Models," Manuscript, Department of Economics, University of California, Berkeley, August, 2002.

Polachek, S. and B. Yoon, "Estimating a Two-Tiered Earnings Function," Working Paper, Department of Economics, State University of New York, Binghamton, 1994.

Polachek, S. and B.Yoon, "Panel Estimates of a Two-Tiered Earnings Frontier," *Journal of Applied Econometrics*, 11, 1996, pp. 169-178.

Prentice, R. and L. Gloeckler., "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data," *Biometrics*, 34, 1978, pp. 57-67.

Rao, C., *Linear Statistical Inference and Its Application*, New York, John Wiley and Sons, 1973.

Rasch, G., "Probabilistic Models for Some Intelligence and Attainment Tests," Copenhagen, Denmark Paedogiska, 1960.

Sueyoshi, G., "Techniques for the Estimation of Maximum Likelihood Models with Large Numbers of Group Effects," Manuscript, Department of Economics, University of California, San Diego, 1993.