A Monte Carlo Study of Bias Corrections for Panel Probit Models

Blair Alexander
Australian National University

Robert Breunig*
Australian National University

December 23, 2013

Abstract

We examine bias corrections which have been proposed for the Fixed Effects Panel Probit model with exogenous regressors, using several different data generating processes to evaluate the performance of the estimators in different situations. We find a best estimator across all cases for coefficient estimates, but when the marginal effects are the quantity of interest no analytical correction is able to outperform the uncorrected maximum likelihood estimator (MLE).

JEL CODES: C13,C23
KEYWORDS: bias correction; panel probit; marginal effects

1 Introduction

Binary response panel data models remain of major interest in microeconometrics. They may be used to address questions not possible with linear models and OLS estimation and are useful for modelling qualitative features of individuals, firms, or countries. Where economic theory makes predictions about labour market participation, employment, educational attainment, marriage status, or health status, binary response models are appropriate, with probit and logit models the most commonly used forms.

Panel data can be used to address issues not feasible with other kinds of data. Where unobserved error terms may be correlated with observable variables, an omitted variable bias will exist which is difficult to address convincingly in a cross-sectional data set. Panel data models allow unobservable terms to be decomposed into a time-constant individual effect, potentially related to the observable variables, and an independent idiosyncratic error. By observing the same individuals across time, it is possible to remove

*Corresponding author: Research School of Economics, Australian National University, Canberra ACT, Australia; E-mail: robert.breunig@anu.edu.au.

the effects of time-constant unobservables and regain consistent estimation. These approaches can also be used to increase the efficiency of an estimator.

A commonly cited example is the effect of education and innate ability on employment or wage: we expect that both education and ability will be associated with better labour market outcomes. However, the two variables are related to each other, with one observed and the other unobserved. If we aim to measure the effect of education keeping ability constant then individual-level panel data is required.

Fixed effects is the commonly preferred estimation method in linear panel models because it is robust to any distribution of observed covariates and time-invariant unobserved effects. Another commonly-employed model, Random Effects, is more efficient if unobserved variables are uncorrelated with observable variables, conditions which will rarely hold in practice.[1]

In the (nonlinear) panel probit model, fixed effects estimation suffers from inconsistency under the *incidental parameters problem*, discussed in section 2.1 below, following a review of the panel probit model itself. Section 3 explains the methodology that will be used, presenting the hypotheses we wish to test and the data generating processes to be used in Monte-Carlo studies. Section 4 gives the results and discussion, while section 5 concludes.

2 Fixed Effects Probit Model

Letting $i \in \{1, ..., N\}$ index an individual in the sample, and $t \in \{1, ..., T\}$ index time, the panel probit model takes the form $y_{it} = 1\left[y_{it}^* > 0\right]$, where y_{it} is a binary response observed, and $y_{it}^* = X_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \, : \, \varepsilon_{it} \sim iid\, N(0,1)$.

X_{it} is a $1 \times k$ vector of observable covariates for individual $i \in \{1, ..., N\}$, at time $t \in \{1, ..., T\}$. X_i is a $T \times k$ matrix formed by stacking the vectors X_{it} across time, and $\bar{X}_i$ is a $1 \times k$ vector of within-individual averages across t. $\boldsymbol{\beta}$ is a k-vector of unknown coefficients for each corresponding covariate in X_{it}.

The variable y_{it}^* is a latent (unobserved) variable assumed to be a linear combination of observable covariates, plus an unobservable error term made up of a time-constant individual effect, α_i, and an idiosyncratic (i.i.d.) error term, ε_{it}, assumed to follow a

[1] Chamberlain (1980) offers a third approach, a Correlated Random Effects (CRE) model which assumes a parametric form of dependence between the two elements.

standard normal distribution in probit models and a logistic distribution in logit models. In the probit model, the restriction of the variance of ε_{it} to 1 gives identification. Some closely related other models take the same form of the main equation, but place different assumptions on the distribution of ε_{it} and α_i. The logit model assumes that ε_{it} terms are independent draws from a standard logistic distribution, leading to a different CDF used in the likelihood function. The Random Effects (RE) and the less commonly-estimated Correlated Random Effects (CRE) probit models place restrictions on the form of dependence between the individual effect α_i and observed covariates X_i. Specifically, independence in the RE case, and a linear form in averages over time for the CRE case: $\alpha_i = \bar{X}_i\delta + \eta_i \, : \, \eta_i \sim N(0, \sigma^2)$.[2]

When estimating by fixed effects (FE), no restrictions are assumed on the distribution of $(\alpha_i|X_i)$. Dummy variables are introduced for each group or individual i, dropping the regression constant as necessary. Since the model is estimated via maximum likelihood, the likelihood contribution from any individual whose response observations do not change throughout the sample can be made arbitrarily close to 1, regardless of the parameter value β.[3] This means that such non-changing observations can be dropped when estimating without affecting the results.

The log-likelihood of a single observation is:

$$
\begin{aligned}
l_{it}(\beta; y_{it}, X_{it}, \alpha_i) &:= lnPr(Y_{it} = y_{it}|X_{it}, \alpha_i) \\
&= ln\left[\Phi(X_{it}\beta + \alpha_i)^{y_{it}}(1 - \Phi(X_{it}\beta + \alpha_i))^{1-y_{it}}\right] \\
&= ln\Phi((2y_{it} - 1)(X_{it}\beta + \alpha_i)) \qquad (1)
\end{aligned}
$$

Where the final line follows from the symmetry of the standard normal CDF: $1 - \Phi(z) = \Phi(-z)$, and the binary nature of y_{it}, allowing the expressing of both one and zero responses in a single index argument in the CDF term. These properties simplify calculations. The total log-likelihood across all observations is:

$$
l(\beta; X, y, \alpha) = \sum_{i=1}^{N}\sum_{t=1}^{T} ln\Phi((2y_{it} - 1)(X_{it}\beta + \alpha_i)) \qquad (2)
$$

[2]See Mundlak (1978) and Chamberlain (1980) and discussion in Wooldridge (2010).

[3]To see this consider the joint likelihood of observations across an individual with all "failures" $y_{it} = 0\,\forall t$, then by allowing α_i to decrease without bound the likelihood of this will converge to 1. Similarly, if $y_{it} = 1\,\forall t$ then the likelihood of these observations will increase to 1 as $\alpha_i \to \infty$.

Fixed effects maximum likelihood estimation maximises the loglikelihood across the $k + N$ free parameters. However it suffers from the *incidental parameters problem*, first identified by Neyman and Scott (1948).

2.1 The incidental parameters problem

Neyman and Scott (1948) consider models where parameters are divided into two types, *incidental parameters* or *nuisance parameters*, which only appear in the likelihood of a finite number of observations, and *structural parameters*[4], which affect the distribution of all observations. Asymptotically, there will only ever be a finite amount of information regarding these incidental, nuisance parameters, thus asymptotic inference is inappropriate. Further, in many nonlinear situations, estimators of the parameters of interest will depend on the nuisance parameters and so will also be affected.

In fixed effects panel models this occurs because a parameter is included for each individual's individual effect α_i for $i \in \{1, \ldots, N\}$, and so each α_i will affect only T observations.

The problem has some possible solutions, however these will affect different kinds of models and distributions of data in different ways. If possible, an information-orthogonal reparametrisation, or a sufficient statistic can be used to express the distribution of the parameter(s) of interest in the model in a form which is conditionally independent of the nuisance parameters, the approach termed "conditional maximum likelihood". For panel logit models with exogenous covariates this does provide a solution. However, for panel probit models no such sufficient statistic is believed to exist[5], so alternative approaches are needed to address the large-sample bias and provide small sample improvements. The problem remains an open question, Lancaster (2000) provides a review and survey of the Incidental Parameters Problem over the decades since Neyman and Scott (1948) was first published. Considering a variety of different linear and nonlinear models, their survey notes solutions where they have been found in the form of a sufficient statistic or a reparametrisation.

The incidental parameters problem persists in the panel probit case because the nuisance parameters - a dummy variable for each individual - cannot be separated from

[4] Also called *common parameters* in Lancaster (2000) to avoid confusion with other econometric uses of the phrase "structural parameters".

[5] See Chamberlain (1980).

estimators of coefficients of interest and the marginal effects. If T is held fixed then the coefficient estimates will not converge asymptotically with large-N. As both N and T increase the increasing number of parameters to estimate means that the coefficients will have an asymptotic bias. This bias is the focus of the correction methods explored in this paper. These analytical estimators form an estimate of the (typically first-order) asymptotic bias as $N, T \to \infty$, and apply the estimate to the small-T sample. The resulting bias-corrected estimators can consistently estimate coefficients as both N and T increase, but their small sample performance is unknown and will be tested in this paper through the use of Monte-Carlo simulations.

2.2 Simulation study and proposed bias corrections

Heckman (1991) was the pioneer in using Monte Carlo simulation techniques to study the bias properties of the fixed effects estimator. Heckman (1991) measured the size of the bias in the fixed effects MLE through Monte Carlo experiments with a simple data generating process for the explanatory variables X_{it}. The explanatory DGP used is a "Nerlove process", after the work of Nerlove (1971) in the context of dynamic linear models. In Heckman's study the response variable is an indicator function driven by this Nerlove process x_{it}, with initial condition as given below:

$$y_{it} = 1\{x_{it}\beta + \alpha_i + \varepsilon_{it} > 0\} \tag{3}$$

$$x_{it} = \frac{1}{10}t + \frac{1}{2}x_{i,t-1} + u_{it} \ : \ u_{it} \sim U(-\frac{1}{2}, \frac{1}{2}) \tag{4}$$

$$x_{i0} = 5 + 10u_{i0} \tag{5}$$

All u_{it} are independent uniformly distributed random variables, and $\alpha_i \sim N(0, \sigma_\tau^2)$, independent of X_i. Heckman performs Monte Carlo experiments by varying the values of β and σ_τ^2, and finds a bias towards zero across all experiments, increasing in absolute size with σ_τ^2.

Greene (2004) repeats Heckman's experiments, but finds the opposite conclusion– that bias is uniformly away from zero.[6] This result also matches that found in fixed effects logit models, where the bias is again away from zero. Greene's main contribution relates to the comparative performance of three different estimation methods - RE,

[6] This result is found again in Hahn and Newey (2004) and Fernandez-Val (2009), suggesting an error on Heckman's part.

Pooled regression, and FE - when all three are known to be inconsistent. To test this he requires a DGP with dependency between X_i and α_i and uses the process:

$$x_{it} \sim iid\, N(0,1) \tag{6}$$

$$\alpha_i = \sqrt{T}\bar{x}_i + \eta_i \,:\, \eta_i \sim N(0,1). \tag{7}$$

The model also includes a discrete explanatory variable d_{it}, with response $y_{it} = 1\{x_{it}\beta + \delta d_{it} + \alpha_i + \varepsilon_{it} > 0\}$, $d_{it} := 1\{x_{it} + h_{it} > 0\} \;:\; h_{it} \sim iid\; N(0,1)$. His key finding is that the Random Effects estimator significantly under-performs when the restrictive assumption of independence between X_i and α_i is false, and $\hat{\beta}_{MLE}$ is again biased away from zero. Qualitatively similar results hold for marginal effects of the discrete variable as for the continuous x_{it}.

Hahn and Newey (2004) propose three methods to correct the first order bias, focusing on fixed effects estimators only. A first method based on a jackknife estimate of the bias, and two analytic expressions estimating the asymptotic first-order bias of the MLE–one based on the Bartlett equations and one based on the M-estimation equations. To test the proposed corrections they use the Nerlove process, enabling comparisons with the earlier work by Greene and Heckman. There is a slight modification of initial condition (5) to $x_{i0} = u_{i0}$.

Fernandez-Val (2009) follows a similar process to Hahn and Newey (2004), proposing an alternative analytical bias correction and testing it with a Monte-Carlo study using the Nerlove process, with initial conditions as in Hahn and Newey (2004). The correction proposed by Fernandez-Val (2009) makes use of more features of the model than the earlier corrections, with adjustments included for dynamic models.

Dhaene and Jochmans (2012) propose a jackknife estimator which combines the maximum likelihood estimator with maximum likelihood estimators based upon splitting the panel by time. In the simplest case, all cross-sectional units and the first $T/2$ consecutive time periods are used to create one estimator and then all cross-sectional units over the last $T/2$ consecutive time periods are used to create a second estimator. The average of the two split-sample estimators is subtracted from two times the maximum likelihood estimator to create the split-panel jackknife estimator.

These five estimators, along with the uncorrected fixed effects maximum likelihood estimator, are considered in this paper.[7] In what follows, we term the estimator proposed by Hahn and Newey (2004) based on the Bartlett equations as "Bartlett", on the M-estimation equation as "HN04-M", and that proposed by Fernandez-Val (2009) as "FV09". We refer to the split-panel jackknife estimator as "SPJ". Formulae for each estimator are reproduced in the appendix.

3 Data Generating Processes

To enable comparison with the literature, the Nerlove process used in Monte-Carlo simulations by Fernandez-Val (2009) has been included. However we believe that a process without a time trend is of more interest as it is easier to compare across alternative DGPs to determine which properties of the process are driving the results. Our process does not include a deterministic time trend and instead adopts a simple AR(1) form for the explanatory variables:

$$x_{it} = \rho x_{it-1} + e_{it} \; : \; e_{it} \sim N(0,1), \quad x_{i0} \sim N(0,1) \tag{8}$$

We will also vary the level of dependence between the observed covariates and the individual effect by defining the average over time for the i-th cross-sectional unit as $\bar{x}_i := \frac{1}{T}\sum_{t=1}^{T} x_{it}$ and setting their individual effect as:

$$\alpha_i = \gamma\sqrt{T}\bar{x}_i + \eta_i \; : \; \eta_i \sim N(0,1). \tag{9}$$

The response variables are determined by

$$y_{it} = 1\left\{x_{it}\beta + \alpha_i + \varepsilon_{it} > 0\right\} \; : \; \varepsilon_{it} \sim N(0,1). \tag{10}$$

We vary the level of dependence between observed covariates and the individual effects by changing γ, while ρ allows us to test different strengths of autocorrelation.

We investigate which of the proposed bias corrections perform best for estimating coefficients and which perform best for estimating average partial effects. We also investigate the following hypotheses:

1. Weaker correlation between the individual effects and the observable covariates will lead to a lower bias in the estimators.

[7] Arellano and Hahn (2006) and Arellano and Bonhomme (2009) propose solutions for the case where T is large (or at least as large as N) so we do not consider these approaches here as we are interested in the case of large N and small T. Bartolucci and Nigro (2012) propose solutions for the logit model.

2. Higher variance of α relative to the variance of x_{it} will lead to a greater bias.
3. Autocorrelation in the x_{it} process will have little effect on the magnitude of the bias, but reduce the precision of estimates.
4. There will be little change in estimation accuracy with increases in the sample size N when T is small and held constant.

In section 4 below we confirm that hypotheses one and two hold. We find evidence against hypothesis three–autocorrelation in x_{it} does increase the bias and the variance of the estimator. With respect to hypothesis 4, we find that while the time dimension is significantly more important, there are still gains to be made from increasing the sample in the N dimension when T is small.

When reporting Average Partial Effects (APEs), we define the quantity of interest to be the marginal effect averaged across the sample:

$$\mu := \frac{1}{NT} \sum_{i,t} \theta \phi(x_{it}\theta + \alpha_i) \tag{11}$$

In section 4 we estimate this as $\tilde{\mu}$ and then calculate the ratio of the estimate to the true value, to find $\frac{\tilde{\mu}}{\mu}$ for each of the appropriate correction methods. The true average partial effect μ will change depending on the sample drawn, however the true value of the ratio ($\frac{\tilde{\mu}}{\mu}$) will always be unity. The jackknife and analytical corrections for $\tilde{\mu}$ are detailed in the appendix, while for the uncorrected MLE ($\hat{\mu}$), the ratio of the estimate to the true value of μ is:

$$\frac{\hat{\mu}}{\mu} = \frac{\frac{1}{NT} \sum_{i,t} \hat{\theta} \phi(x_{it}\hat{\theta} + \hat{\alpha}_i)}{\frac{1}{NT} \sum_{i,t} \theta \phi(x_{it}\theta + \alpha_i)} \tag{12}$$

where $\hat{\theta}, \hat{\alpha}_i$ are the MLEs for the coefficient parameter and the individual effects respectively. The contribution to the sum in the numerator from a non-changing individual will be zero since $\hat{\alpha}_i$ is unbounded.

3.1 Experiments

The experiments we use to test the hypotheses are based around an AR(1) process for x_{it}, as defined in equations (8) and (9). Prior to that however, our first experiment reproduces the Monte Carlo tests of Fernandez-Val (2009) by using a Nerlove process as in equation (4), with initial condition $x_{i0} = u_{i0}$, panel lengths $T = 4$, 8, and sample size $N = 100$. This enables a comparison with the earlier literature, and with the other experiments using an AR(1) process for x_{it}.

Experiment 2 adopts an iid process for x_{it}. This uses the process defined by equations (8) and (9) with $\rho = 0$ and $\gamma \in \{-1, 0, 1\}$, setting $N = 100$ and $T = 8$. This process is similar to the experiments of Greene (2004), but excludes the discrete covariate from the process. The data generating process itself is $x_{it} \sim iid\ N(0, 1)$, $\alpha_i := \frac{1}{\sqrt{T}} \sum_t x_{it} + \eta_i$: $\eta_i \sim N(0, 1)$.

Experiment 3 repeats experiment 2 but sets $\rho = 0.8$ instead to test the effect of strong positive autocorrelation in the data generating process on the estimators. Again we consider each case in $\gamma \in \{-1, 0, 1\}$. Comparing the results of this experiment with those of experiment 2 is a test of hypothesis 3–that autocorrelation in x_{it} will not affect the bias properties of the estimator.

Experiment 4 is designed to complement the middle case of experiment 2. Since the total variance of α_i is smaller when $\gamma = 0$, we use $\alpha_i = \sqrt{2}\eta_i$: $\eta_i \sim N(0, 1)$ to maintain constant variance of the individual effects. It is also compared with experiment 2 to test the hypothesis that higher variance of α_i relative to that of x_{it} will lead to greater bias.

Finally, experiment 5 explores how sample size effects the precision of estimates by repeating two earlier cases with $N = 400$–the Nerlove process from 1 with $T = 4$, and experiment 2's test of $(\gamma, \rho) = (0, 0)$ with $T = 8$.

4 Results and Discussion

We now report the results of the experiments. Firstly, section 4.1 covers experiment 1, as performed by Fernandez-Val (2009). We also present a kernel density plot of the sampling distributions for the $T = 4$ case, the sampling distribution of coefficients in figure 1 and average partial effects in figure 2. In section 4.2 we present the results concerning estimating coefficients from experiments 2 and 3 in tables 3 and 4 respectively, along with a plot of the sampling distribution for the case where $(\rho, \gamma) = (0, 1)$, and discuss the results in the context of hypotheses 1-3 for coefficients. Section 4.3 considers the corresponding APEs from experiments 2 and 3 and relates results to bias properties of the APEs. Finally, section 4.4 reports results from increasing the sample "width" N without increasing length T, and relates the results from both this test and the Nerlove process with increased sample size in section 4.1 to hypothesis 4.

4.1 Nerlove Process

As found by Fernandez-Val (2009), his proposed estimator clearly performs the best of the set in estimating coefficients, as it makes use of the known probit structure to improve over more generally applicable approaches such as the jackknife. Table 1 gives summary statistics of the experiments with $T = 4$ and $T = 8$. In Figure 1 we also plot a kernel density estimate of the five estimators under comparison for the case where $T = 4$. This allows us to present more information than only the summary statistics that were reported in the previous work. The distribution of the MLE estimator is positively skewed, which is reduced by either the jackknife or FV09 bias corrections, but amplified by either of the estimators proposed by Hahn and Newey (2004). The split-panel jackknife performs very poorly, particularly when T is small.

Table 1
Summary statistics for Nerlove Processes, coefficients

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	1.402	0.755	1.105	1.211	1.054	0.580
$T = 4$	median	1.376	0.760	1.084	1.199	1.041	0.631
	st. dev.	0.404	0.253	0.315	0.340	0.284	0.823
	rmse	0.570	0.352	0.332	0.400	0.289	0.924
	$\overline{x}$	1.188	0.957	1.058	1.098	1.023	0.942
$T = 8$	median	1.191	0.958	1.060	1.100	1.025	0.944
	st. dev.	0.146	0.114	0.129	0.127	0.120	0.251
	rmse	0.238	0.122	0.142	0.160	0.123	0.257
$T = 4$	$\overline{x}$	1.402	0.800	1.110	1.209	1.058	0.643
	median	1.400	0.800	1.107	1.207	1.058	0.671
$N = 400$	st. dev.	0.202	0.129	0.160	0.168	0.142	0.358
	rmse	0.449	0.238	0.194	0.268	0.154	0.505

$N = 100$ in the top two panels. Third panel included for comparison with section 4.4 below.
Standard deviation and root mean squared error are abbreviated to *st. dev.* and *rmse*, respectively.

There are significant improvements demonstrated by increasing the panel length T, with both standard error and absolute bias decreasing more than proportionally with T. We also see that the FV09 estimator outperforms all others in terms of bias and mean-squared error. The jackknife estimator typically increases the variance of estimates; while in this case it is actually producing the lowest standard errors, this is not a result that generalizes to APEs.

Table 2 shows the results for the ratio of average partial effects to the true value, while figure 2 shows the estimated sampling distribution for the APEs for the "short panel" case of $T = 4$, complementing the density of coefficient estimates previously. As for the coefficients, there is a significant increase in precision with the increased panel length $T = 8$, over twice that for the shorter panel.

Table 2

Summary statistics for Nerlove Processes, APE ratios

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	0.985	1.016	0.975	1.055	0.928	1.050
$T = 4$	median	0.986	1.021	0.978	1.059	0.930	1.054
	st. dev.	0.253	0.285	0.257	0.269	0.234	0.407
	rmse	0.253	0.285	0.259	0.274	0.245	0.410
	$\overline{x}$	0.998	1.007	1.007	1.035	0.983	1.002
$T = 8$	median	0.994	1.004	1.003	1.031	0.979	0.993
	st. dev.	0.106	0.109	0.109	0.106	0.104	0.178
	rmse	0.106	0.110	0.109	0.112	0.106	0.178
$T = 4$	$\overline{x}$	0.995	1.027	0.987	1.063	0.938	1.038
	median	0.992	1.021	0.983	1.062	0.936	1.046
$N = 400$	st. dev.	0.128	0.144	0.132	0.135	0.118	0.198
	rmse	0.128	0.147	0.133	0.150	0.133	0.202

Figure 1: Coefficient Estimates, Nerlove Process, $T = 4$

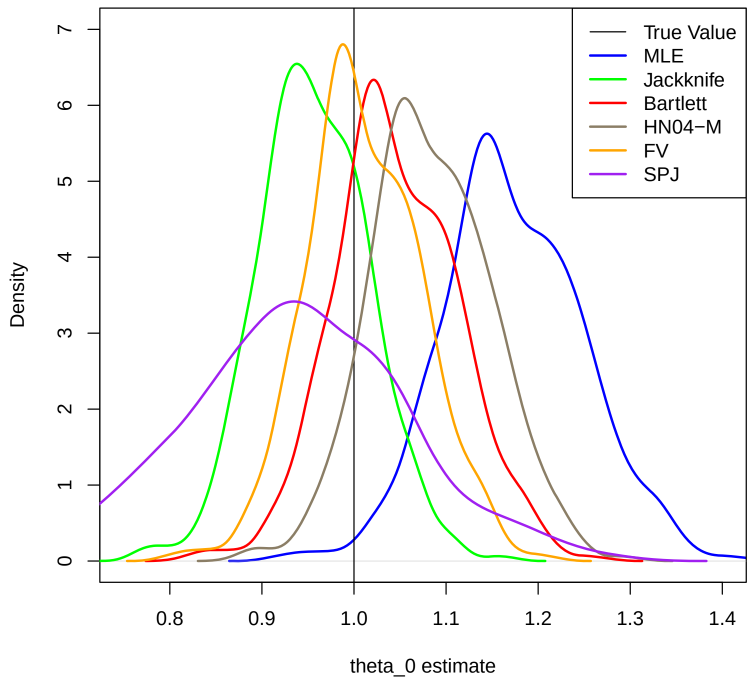


Figure 2: APE ratio, Nerlove Process, $T = 4$

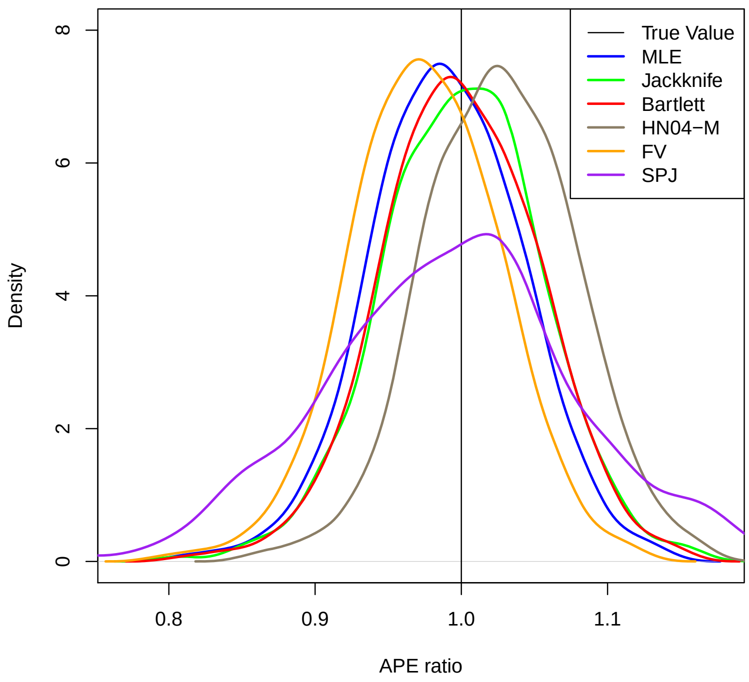


Figure 2 shows that there is no clear best estimator from the five considered–all have very little bias, with variance approximately similar. Only the jackknife stands out as having a slightly wider standard deviation. These results align with the "small bias property" reported in Fernandez-Val (2009), where estimated asymptotic bias of marginal effects was estimated to be typically less than 1% of the true value, though the extent that this extends to more complicated data generating processes with multiple variables has not yet been explored. While the differences are small, no estimator outperforms the uncorrected MLE; only in one particular simulation did the Bartlett-based bias correction perform as well as the MLE.

4.2 Coefficients under AR process

Tables 3 and 4 give the results for differing values of (ρ, γ). Throughout we have used a simulated sample size of $N = 100$, length $T = 8$, with $B = 400$ simulation runs.[8]

[8] Unreported experiments found large outliers occurred when panel length was shorter, e.g., $T = 4$, which were then typically amplified by the analytical corrections being considered. This suggests FE estimators should not be used in such situations because of a significant lack of within-individual data.

As in the previous section, we also present a kernel density estimate of the estimators considered. This sampling distribution for the model with $(\rho, \gamma) = (0, 1)$ is presented in figure 3.

Table 3

Summary statistics for uncorrelated process $(\rho = 0)$*, coefficients*

Parameters	Statistic	FE MLE	Jack-knife	Bartlett	HN04 -M	FV09	SPJ
	$\overline{x}$	1.234	0.934	1.065	1.053	0.999	0.890
$\gamma = 1$	median	1.227	0.925	1.058	1.047	0.994	0.889
	st. dev.	0.133	0.093	0.105	0.110	0.091	0.155
	rmse	0.269	0.113	0.123	0.122	0.091	0.190
	$\overline{x}$	1.214	0.932	1.061	1.038	0.998	0.883
$\gamma = 0$	median	1.211	0.931	1.062	1.039	0.997	0.880
	st. dev.	0.110	0.074	0.089	0.090	0.078	0.113
	rmse	0.241	0.101	0.108	0.098	0.078	0.163
	$\overline{x}$	1.214	0.931	1.058	1.015	0.995	0.876
$\gamma = -1$	median	1.202	0.927	1.050	1.010	0.990	0.879
	st. dev.	0.119	0.081	0.094	0.098	0.084	0.139
	rmse	0.245	0.107	0.110	0.099	0.084	0.186

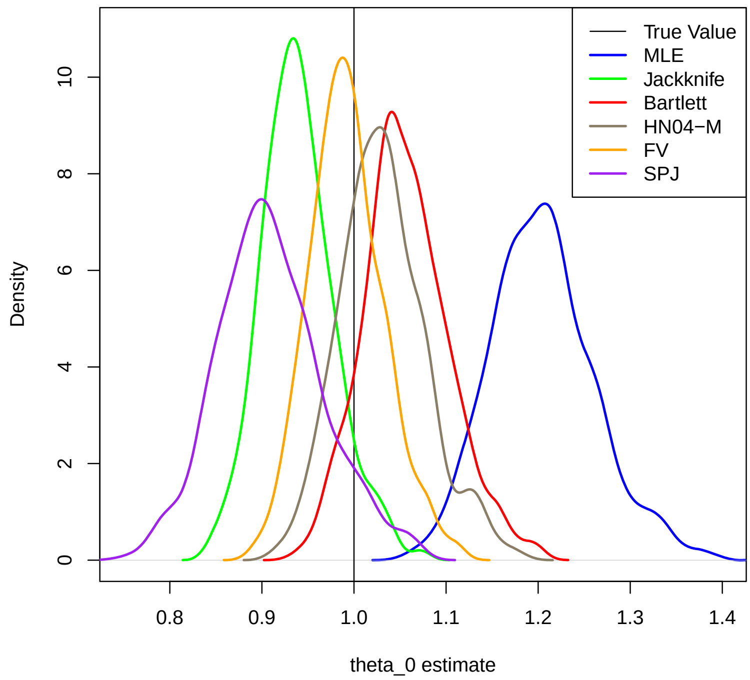
Table 4

Summary statistics for autocorrelated process $(\rho = 0.8)$*, coefficients*

Parameters	Statistic	FE MLE	Jack-knife	Bartlett	HN04 -M	FV09	SPJ
	$\overline{x}$	1.300	0.920	1.079	1.115	1.013	0.901
$\gamma = 1$	median	1.271	0.909	1.069	1.087	1.007	0.921
	st. dev.	0.225	0.153	0.156	0.191	0.139	0.378
	rmse	0.375	0.173	0.175	0.223	0.140	0.391
	$\overline{x}$	1.243	0.929	1.061	1.103	0.994	0.925
$\gamma = 0$	median	1.232	0.921	1.054	1.088	0.984	0.926
	st. dev.	0.137	0.095	0.105	0.121	0.092	0.169
	rmse	0.279	0.118	0.121	0.159	0.092	0.185
	$\overline{x}$	1.258	0.927	1.073	0.989	1.003	0.937
$\gamma = -1$	median	1.255	0.925	1.066	0.988	1.003	0.944
	st. dev.	0.146	0.098	0.116	0.116	0.099	0.194
	rmse	0.296	0.122	0.137	0.117	0.099	0.204

The results are qualitatively similar to those for the Nerlove process in the previous section. Across all tests shown in tables 3 and 4 the jackknife estimate of the bias is too high, so the jackknife-corrected estimator over-corrects on average giving a bias-corrected value that is too low. The correction proposed by Fernandez-Val outperforms all others in every case in terms of the mean-squared error, with lowest absolute size of

Figure 3: Coefficient Estimates, $(\rho, \gamma) = (0, 1)$, $T = 8$



the bias and low standard error of the estimator, suggesting that it is the best to use for estimating the coefficients.

The results also suggest that the direction of correlation has only a small impact on the bias, except where there is both strong autocorrelation in the observed process x_{it} and correlation between the individual effect α_i and $\bar{x}_i$. Unfortunately, this may often be the case in empirical practice; for example–income or wealth level may be a predictor of health status, and would be both highly autocorrelated and positively correlated with other unobserved individual effects. Precise estimation also becomes quite difficult in these cases, since effective sample size shrinks rapidly with persistent x_{it} causing y_{it} to change for fewer individuals.

In relation to hypothesis 1, we confirm that uncorrelated α_i and x_{it} are beneficial for estimation, but the effects are quite small and dominated by other features of the data. The bias was found to be strongest when correlation was positive, with only a small difference between when $\gamma = 0$ and $\gamma = -1$.

Further, contrary to our third hypothesis, we found that there was a significant increase in bias when x_{it} was set to be autocorrelated. This occurred across all dependency values for γ. There was also an increase in the variance of estimates, driven partly by data loss as y_{it} would change for fewer members of the sample. The implication for researchers is to look for higher "within" variation in the covariates, similar to evaluating treatment effects.

Finally, our test with higher variance for α_i is given in table 5. The results support hypothesis 2, and agree with Greene's (2004) findings on the effects that variance of unobserved effects has on the bias of the estimator. The same result that Greene found for the MLE also holds across all of the estimators considered in this paper–that higher variance in α leads to higher bias.

Table 5

Summary statistics for high-variance α process, $\alpha \sim N(0, 2)$, coefficients

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	1.264	0.905	1.044	1.080	0.964	0.874
$\gamma = 0$	median	1.254	0.900	1.037	1.074	0.963	0.876
	st. dev.	0.115	0.075	0.078	0.099	0.062	0.156
	rmse	0.288	0.121	0.090	0.128	0.072	0.201

4.3 Marginal Effects

Agreeing with the results of Fernandez-Val (2009) regarding the bias of APEs in a Nerlove process, we found little difference in this bias across estimators. However, the Nerlove process generates individual effects independently of the observable covariates x_{it}. The main reason to use fixed effects estimation is for problems where there is an unknown type of dependence between observable X_{it} and unobservable α_i, without assuming a specific form for the dependence.

The earlier results of section 4.2 using modifications of Greene's DGP showed relatively small changes in the coefficient estimates, but these results do not hold for marginal effects. When dependency is introduced the bias in the marginal effects increases with dependency, and is amplified by positive correlation, $\gamma = 1$.

The uncorrected MLE of average partial effects typically underestimates the true value since it assigns a zero contribution to the average effect of any individual who does not change state over the sample period. Whereas the true value must always be strictly positive, the MLE fits zero to marginal effect $\hat{\theta}\phi(x_{it}\hat{\theta}+\hat{\alpha}_i)$ because the individual term $\hat{\alpha}_i$ is unboundedly large. However, the analytical bias corrected estimates of the APE were below the MLE on average (except for the case of the Nerlove process of Table 2). In some cases the jackknife was able to improve the average bias performance of the ratio of estimated to true APE, but at the cost of increased standard error of estimates. There is no overall gain in terms of mean squared error.

Table 6
Summary statistics for uncorrelated process ($\rho = 0$), APE ratios

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	0.985	1.004	0.959	0.951	0.959	1.024
$\gamma = 1$	median	0.983	1.003	0.962	0.953	0.958	1.022
	st. dev.	0.065	0.069	0.064	0.065	0.061	0.088
	rmse	0.066	0.070	0.071	0.081	0.073	0.091
	$\overline{x}$	0.998	1.005	0.989	0.979	0.976	1.023
$\gamma = 0$	median	1.004	1.007	0.991	0.982	0.981	1.021
	st. dev.	0.052	0.053	0.052	0.051	0.049	0.069
	rmse	0.052	0.053	0.053	0.055	0.055	0.072
	$\overline{x}$	0.994	1.004	0.981	0.963	0.971	1.022
$\gamma = -1$	median	0.989	1.000	0.977	0.958	0.967	1.017
	st. dev.	0.057	0.059	0.055	0.055	0.054	0.077
	rmse	0.057	0.059	0.058	0.066	0.061	0.080

Table 7
Summary statistics for highly autocorrelated process ($\rho = 0.8$), APE ratios

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	0.977	1.005	0.910	0.883	0.940	1.011
$\gamma = 1$	median	0.972	1.000	0.918	0.915	0.941	1.000
	st. dev.	0.113	0.128	0.138	0.242	0.105	0.158
	rmse	0.115	0.128	0.165	0.269	0.121	0.158
	$\overline{x}$	0.996	1.010	0.952	0.954	0.964	1.029
$\gamma = 0$	median	0.995	1.010	0.956	0.964	0.965	1.031
	st. dev.	0.067	0.074	0.072	0.088	0.062	0.096
	rmse	0.067	0.074	0.087	0.099	0.072	0.101
	$\overline{x}$	0.984	1.005	0.948	0.916	0.953	1.025
$\gamma = -1$	median	0.983	1.004	0.947	0.914	0.952	1.029
	st. dev.	0.075	0.083	0.075	0.072	0.070	0.106
	rmse	0.077	0.083	0.091	0.110	0.085	0.109

Tables 6 and 7 give the results for the equivalent APE ratio for each of the tests in tables 3 and 4 earlier. Figure 4 shows the estimated sampling distribution for the $(\rho, \gamma) = (0, 1)$ case. The small negative bias on the MLE decreases when $\gamma = -1$, but is lowest when $\gamma = 0$ and α_i and x_{it} are independent.

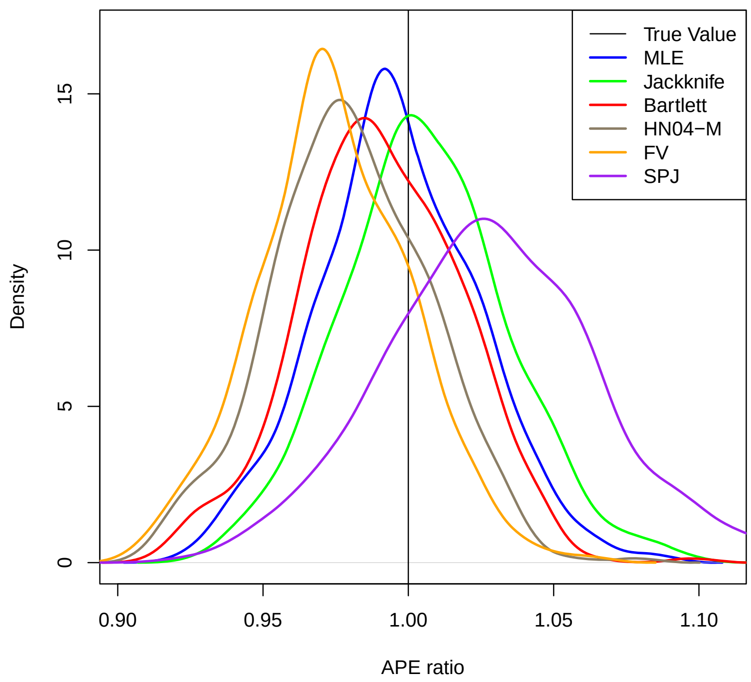
Under experiment 4 using the process $x_{it} \sim iid\ N(0, 1)$ and $\alpha_i \sim N(0, 2)$, we found both the Bartlett and HN04-M estimators were subject to highly influential outliers that gave a strong negative skew to the estimators and significantly impaired their performance. Table 8 gives the results for APE ratios across the different estimators for this experiment.

Table 8

Summary statistics for high-variance α process, $\alpha \sim N(0, 2)$, APEs

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	1.002	1.012	0.864	0.775	0.965	1.065
$\gamma = 0$	median	1.000	1.009	0.898	0.889	0.965	1.063
	st. dev.	0.042	0.047	0.180	0.707	0.037	0.068
	rmse	0.042	0.049	0.225	0.742	0.051	0.094

Figure 4: APE ratio, $(\rho, \gamma) = (0, 1)$, $T = 8$



The MLE of the average partial effects was found to be correct on average, and so no improvements were possible. The FV09 estimator was not subject to extreme outliers, but still under-performs both the MLE and Jackknife approaches. In relation to hypothesis 2 we conclude that when considering marginal effects the two estimators proposed by Hahn and Newey (2004) are very sensitive to variance in the individual effects and should not be used; while the impact of increasing the variance of α on the MLE's properties was negligible.

4.4 Sample Size Effects

The results from experiment $x_{it}, \alpha_i \sim iid\ N(0, 1)$, with $N = 400$ and $T = 8$ are given in tables 9 and 10 below. The test using a larger sample with a Nerlove process is in the bottom panel of table 1 for easy comparison with the experiment with $N = 100$. We find that there is no change in the bias of the estimators under either experiment, but that precision is about twice what it was with $N = 100$ for both coefficients and APEs.

While increased precision of an asymptotically biased estimator (the MLE for coefficients) is of little value of itself, the bias corrections for the coefficients also benefit from the increased precision. In response to hypothesis 4, we find that while the time dimension is significantly more important, there are still gains to be made from a broader sample. A doubling of panel length approximately doubled precision and reduced the bias of all estimates, while a quadrupling of panel width doubled the precision only.

Table 9

Summary statistics for $(\rho, \gamma) = (0, 0)$ process, large N, coefficients

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	1.205	0.939	1.057	1.031	0.992	0.909
$(\rho, \gamma) = (0, 0)$	median	1.203	0.936	1.052	1.029	0.990	0.905
	st. dev.	0.056	0.039	0.046	0.046	0.040	0.057
	rmse	0.212	0.072	0.073	0.056	0.041	0.107

Table 10

Summary statistics for $(\rho, \gamma) = (0, 0)$ process, large N, APE ratios

Parameters	**Statistic**	**FE MLE**	**Jack-knife**	**Bartlett**	**HN04 -M**	**FV09**	**SPJ**
	$\overline{x}$	0.997	1.006	0.990	0.980	0.975	1.028
$(\rho, \gamma) = (0, 0)$	median	0.995	1.004	0.989	0.979	0.973	1.027
	st. dev.	0.028	0.029	0.027	0.027	0.026	0.037
	rmse	0.028	0.029	0.029	0.034	0.037	0.046

5 Conclusion

Throughout this paper we have focussed on only a single covariate and associated parameter of interest, whereas in practice models typically have several, perhaps dozens, of regressors, both continuous and discrete. How the potential interactions among multiple observable regressors affects the numerical performance of the estimators remains unexplored, albeit difficult to ascribe to a single source.

In a model with a single predictor variable we may only be interested in the sign of the coefficient–whether the variable has an increasing or decreasing effect–since the coefficients have no direct interpretation. In this case we know that the MLE will be biased away from zero, and so can effectively increase the power of inference by using a biased estimator. This is a very particular case, and would not generalise to more detailed models and problems of inference.

Where the quantity of interest is the marginal effect of a variable, the results of this paper, along with the earlier work which tested only Nerlove processes specifically, suggest that the MLE or jackknife bias corrections are superior to the three analytical proposals made in Hahn and Newey (2004) and Fernandez-Val (2009) and to the split-sample jackknife of Dhaene and Jochmans (2012). Whether a new analytical estimator focussed on the marginal effects specifically can be developed is an area for future research.

References

Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models, *Econometrica* **77**(2): 489–536.

Arellano, M. and Hahn, J. (2006). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. CEMFI Working Paper Number wp2006_0613. Availabe at `http://www.cemfi.es/ftp/wp/0613.pdf`.

Bartolucci, F. and Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data, *Journal of Econometrics* **170**(1): 102–116.

Chamberlain, G. (1980). Analysis of covariance with qualitative data, *Review of Economic Studies* **47**(146): 225–238.

Dhaene, G. and Jochmans, K. (2012). Split-panel jackknife estimation of fixed-effects models. Available at `http://www.econ.kuleuven.be/ew/academic/econmetr/members/Dhaene/Papers/dhaene-jochmans%20jackknife.pdf`.

Fernandez-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models, *Journal of Econometrics* **150**(1): 71–86.

Greene, W. (2004). The behaviour of the maximum likelihood estimator in limited dependent variable models in the presence of fixed effects, *The Econometrics Journal* **7**(1): 98–119.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models, *Econometrica* **72**(4): 1295–1319.

Heckman, J. J. (1991). 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. L. McFadden (eds), *Structural Analysis of Discrete Data and Econometric Applications*, Cambridge, MA: MIT Press, pp. 179–195.

Lancaster, T. (2000). The incidental parameter problem since 1948, *Journal of Econometrics* **95**: 391–413.

Mundlak, Y. (1978). On the pooling of time series and cross section data, *Econometrica* **46**: 69–85.

Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections, *Econometrica* **39**(2).

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations, *Econometrica* **16**(1).

Quenouille, M. H. (1956). Notes on bias in estimation, *Biometrika* **43**: 353–360.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples (abstract), *Ann. Math. Statist.* **29**: 614.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, second edn, The MIT Press, Cambridge, MA.

Appendix

Bias correction equations

We call the three analytical bias corrections which are considered "Bartlett", "HN04-M" and "FV09". As we focus on the case of the probit model with only a single exogenous regressor, the terms involving dynamics can be dropped out of the equation, the relevant pdf will be known to be the standard normal, and all vectors of coefficients X_{it} will actually be scalars, so the transpose will equal the original vector: $x_{it}x'_{it} = x^2_{it}$, and the coefficient parameter of interest is also a scalar.

Letting $\hat{\theta}$ be the MLE of the parameter of interest; and $\hat{y}^*_{it} := x_{it}\hat{\theta} + \hat{\alpha}_i$, and $\hat{\Phi}_{it} := \Phi\left((2y_{it} - 1) \times \hat{y}^*_{it}\right)$, $\hat{\phi}_{it} := \phi\left(\hat{y}^*_{it}\right)$ be the estimates of the latent variable, the likelihood of the (i, t)-th observation using the standard normal CDF, Φ, and the standard normal pdf, ϕ, evaluated at the estimated latent variable, respectively.

BC1: Bartlett

The bias correction proposed in Hahn and Newey (2004) based on the Bartlett equations is:

$$\tilde{\theta}_B := \hat{\theta} - \bar{B}(\hat{\theta})$$

Where

$$\begin{aligned}
\bar{B}(\theta) \quad &= \quad \left[\frac{1}{NT}\sum_i\sum_t \frac{\hat{\phi}^2_{it}}{\hat{\Phi}^2_{it}}\left(x_{it} - \sum_t \frac{\hat{\phi}^2_{it}}{\hat{\Phi}^2_{it}}x_{it}\right)\right]^{-1} \\
&\quad \times \left[\frac{-1}{2N}\sum_i\left(\left[\sum_t \frac{\hat{\phi}^2_{it}}{\hat{\Phi}^2_{it}}\right]^{-1}\sum_t\left[\hat{y}^*_{it}\frac{\hat{\phi}^2_{it}}{\hat{\Phi}^2_{it}}\left(x_{it} - \sum_t \frac{\hat{\phi}^2_{it}}{\hat{\Phi}^2_{it}}x_{it}\right)\right]\right)\right]
\end{aligned}$$

BC2: HN04-M

Hahn & Newey's correction based on the analytic M-estimation is significantly more complicated, it is stated below with little modification:

$$\tilde{\theta}_M = \hat{\theta} - \hat{H}(\theta)^{-1}\hat{b}(\theta)$$

With the components as follows:

$$\begin{aligned}
\hat{b}(\theta) &:= \frac{1}{NT}\sum_i\sum_t\left[\hat{v}_{it\alpha}\times\left(\hat{\beta}_i+\hat{\psi}_{it}\right)+\hat{v}_{it\alpha\alpha}\frac{1}{2}\hat{\sigma}_i^2\right]\times x_{it}\\
\hat{H}(\theta) &:= \frac{1}{NT}\sum_i\sum_t\left[\hat{v}_{it\alpha}x_{it}^2-\hat{v}_{it\alpha}x_{it}\frac{\sum_s\hat{v}_{is\alpha}x_{is}}{\sum_s\hat{v}_{is\alpha}}\right]\\
\hat{\beta}_i &:= -\left[\sum_t\hat{v}_{it\alpha}\right]^{-1}\sum_t\left(v_{it\alpha}\hat{\psi}_{it}+\hat{v}_{it\alpha\alpha}\frac{1}{2}\hat{\sigma}_i^2\right)\\
\hat{\sigma}_i^2 &:= \frac{\frac{1}{T}\sum_t\hat{v}_{it}^2}{\left(\frac{1}{T}\sum_t\hat{v}_{it\alpha}\right)^2}\\
\psi_{it} &:= \frac{-\hat{v}_{it}}{\frac{1}{T}\sum_s\hat{v}_{is\alpha}}\\
\hat{v}_{it} &:= \frac{\partial}{\partial\alpha_i}lnl_{it}=(2y_{it}-1)\frac{\hat{\phi}_{it}}{\hat{\Phi}_{it}}\\
\hat{v}_{it\alpha} &:= \frac{\partial^2}{\partial\alpha_i^2}lnl_{it}=\frac{\hat{\phi}_{it}\hat{\Phi}_{it}-\hat{\phi}_{it}^2}{\hat{\Phi}_{it}^2}\\
\hat{v}_{it\alpha\alpha} &:= \frac{\partial^3}{\partial\alpha_i^3}lnl_{it}=-\hat{v}_{it}-\hat{y}_{it}^*\hat{v}_{it\alpha}-2\hat{v}_{it}\times\hat{v}_{it\alpha}
\end{aligned}$$

Where $l_{it}=\Phi\left((2y_{it}-1)(x_{it}\hat{\theta}+\hat{\alpha}_i)\right)$ is the likelihood contribution of observation it evaluated at the MLE.

BC3: FV09

The FV09 estimator is given by:

$$\tilde{\theta}_V := \hat{\theta}-\frac{\frac{1}{N}\sum_i b_i(\hat{\theta})}{\frac{1}{N}\sum_i I_i(\hat{\theta})}$$

Where:

$$\begin{aligned}
b_i(\theta) &:= \frac{\left(\frac{1}{T}\sum_t\frac{\hat{\phi}_{it}^2x_{it}}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}\right)\left(\frac{-1}{T}\sum_t\frac{\hat{\phi}_{it}^2}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}(2y_{it}-1)\hat{y}_{it}^*\right)}{2\times\left(\frac{1}{T}\sum_t\frac{\hat{\phi}_{it}^2}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}\right)^2}\\
&\quad-\frac{\left(\frac{-1}{T}\sum_t\frac{\hat{\phi}_{it}^2}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}(2y_{it}-1)\hat{y}_{it}^*\times x_{it}\right)}{2\times\left(\frac{1}{T}\sum_t\frac{\hat{\phi}_{it}^2}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}\right)}\\
I_i(\theta) &:= \frac{1}{T}\sum_t\frac{\hat{\phi}_{it}^2x_{it}^2}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}-\left[\frac{1}{T}\sum_t\frac{\hat{\phi}_{it}}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}\right]^{-1}\left(\frac{1}{T}\sum_t\frac{\hat{\phi}_{it}x_{it}}{\hat{\Phi}_{it}(1-\hat{\Phi}_{it})}\right)^2
\end{aligned}$$

Jackknife

The Quenouille-Tukey jackknife[9] is a general estimation technique represented in this context as:

$$\tilde{\theta}_{JK} = T\hat{\theta} - \frac{T-1}{T}\sum_t \hat{\theta}_{(t)}$$

Where $\hat{\theta}_{(t)}$ is the MLE estimated with the t-th time period omitted.

To estimate the average partial effect with a bias correction using the jackknife we use

$$\begin{aligned}
\tilde{\mu}_{JK} &= T \times \left[\frac{1}{NT}\sum_{i,t}\hat{\theta}\phi(x_{it}\hat{\theta}+\hat{\alpha}_i)\right] - \frac{T-1}{T}\sum_t\left[\frac{1}{N(T-1)}\sum_{i,s}\hat{\theta}_{(t)}\phi(x_{is}\hat{\theta}_{(t)}+\hat{\alpha}_{i(t)})\right] \\
&= \frac{1}{N}\sum_{i,t}\hat{\theta}\phi(x_{it}\hat{\theta}+\hat{\alpha}_i) - \frac{1}{NT}\sum_t\sum_{i,s}\hat{\theta}_{(t)}\phi(x_{is}\hat{\theta}_{(t)}+\hat{\alpha}_{i(t)})
\end{aligned}$$

Where $\hat{\theta}$ is the MLE using the complete sample, and $\hat{\theta}_{(t)}$ is the MLE using the subsample with the t-th period removed.

Split-panel Jackknife

The split panel jackknife of Dhaene and Jochmans (2012) operates similarly to the full jackknife but drops half of the panel at a time, making only two resamples needed, saving computation time.

$$\tilde{\theta}_{SPJ} = 2\hat{\theta} - \frac{1}{2}\left(\hat{\theta}_{(1:\frac{T}{2})} + \hat{\theta}_{(\frac{T}{2}+1:T)}\right) \tag{13}$$

and similarly,

$$\tilde{\mu}_{SPJ} = 2\hat{\mu} - \frac{1}{2}\left(\hat{\mu}_{(1:\frac{T}{2})} + \hat{\mu}_{(\frac{T}{2}+1:T)}\right) \tag{14}$$

is the SPJ estimator of the average partial effects.

Where $\hat{\theta}_{(1:\frac{T}{2})}$ and $\hat{\theta}_{(\frac{T}{2}+1:T)}$ are the MLEs based on the appropriate half-panel.

Analytical bias correction to average partial effects

The estimators used by Hahn and Newey (2004) are based on applying the Bartlett equations to the marginal effects. In this case they become

$$\tilde{\mu} := \frac{1}{NT}\sum_i\sum_t\left[\hat{\theta}\phi\left(x_{it}\hat{\theta}+\hat{\alpha}_i\right) - \hat{\theta}\phi'\left(x_{it}\hat{\theta}+\hat{\alpha}_i\right)\frac{\bar{\beta}_i(\theta)}{T} - \hat{\theta}\phi''\left(x_{it}\hat{\theta}+\hat{\alpha}_i\right)\frac{\bar{\sigma}^2(\hat{\theta})}{2T}\right]$$

[9] The jackknife was first introduced by Quenouille (1956) for bias reduction and by Tukey (1958) for variance estimation.

Where:

$$\bar{\beta}_i(\hat{\theta}) := \frac{-\bar{\sigma}_i^2(\hat{\theta})^2}{2T} \sum_t \frac{\hat{\phi}_{it}\hat{\phi'}_{it}}{\hat{\Phi}_{it}^2}$$

$$\bar{\sigma}_i^2(\hat{\theta}) := \frac{T}{\sum_t \left(\frac{\hat{\phi}_{it}}{\hat{\Phi}_{it}}\right)^2}$$

As before, let $\hat{\phi}_{it} := \phi(x_{it}\hat{\theta} + \hat{\alpha}_i)$ and $\hat{\phi'}_{it} := \phi'\Big((2y_{it} - 1)(x_{it}\hat{\theta} + \hat{\alpha}_i)\Big)$ represent the standard normal pdf evaluated at the estimated latent variable for the (i, t)-th observation, and the derivative evaluated at (i, t)-th observation. Because of symmetry the term $(2y_{it} - 1)$ cancels from the first but not the second.

The equations may be evaluated at either $\hat{\theta} = \hat{\theta}_M$ or $\hat{\theta} = \hat{\theta}_B$ to form the appropriate correction.

Fernandez-Val uses a very similar correction, $\tilde{\mu}$'s definition remains the same but the definitions of $\bar{\beta}_i(\theta)$ and $\bar{\sigma}_i^2(\theta)$ are modified to:

$$\hat{\beta}_i(\hat{\theta}) := \frac{-\hat{\sigma}_i^2(\hat{\theta})^2}{2T} \sum_t \frac{\hat{\phi}_{it}\phi'(x_{it}\hat{\theta} + \hat{\alpha}_i)}{(1 - \hat{\Phi}_{it})\hat{\Phi}_{it}}$$

$$\hat{\sigma}_i^2(\hat{\theta}) := \frac{T}{\sum_t \frac{\hat{\phi}_{it}^2}{\hat{\Phi}_{it}(1 - \hat{\Phi}_{it})}}$$

These are then evaluated at $\hat{\theta} = \hat{\theta}_V$ and substituted into the equation above, replacing $\bar{\beta}_i(\theta)$ and $\bar{\sigma}_i^2(\theta)$ respectively. The key differences are that $\hat{\Phi}_{it}(1 - \hat{\Phi}_{it})$ appears in the denominator instead of $\hat{\Phi}_{it}^2$, and that the derivative of the pdf does not "sign" its argument depending on y_{it}: $\phi'(x_{it}\hat{\theta} + \hat{\alpha}_i) \neq \phi'\Big((2y_{it} - 1)(x_{it}\hat{\theta} + \hat{\alpha}_i)\Big)$.