Bias and Inconsistency in Binary Dependent Variable Models

Monte-Carlo-Simulation-Based Performance Analysis

Maastricht University
Bachelor Thesis in Econometrics & Operations Research
Name: Kacper Solarski
Student ID: i6153241
Supervisor: Martin Schumann
Date submitted: September 21, 2021

Abstract

I investigate the most common problems encountered in binary dependent variable models and proposed solutions. In particular, I examine the bias and inconsistency of the following estimators: linear probability model, non-linear fixed effects, and non-linear random effects. I discuss the bias correction for the non-linear fixed effects model proposed by Fernández-Val (2009) and a solution to the initial conditions problem proposed by Wooldridge (2005). Consequently, I conduct a Monte Carlo simulation to compare the performance of various estimation techniques. The simulation indicates that for the estimations of the coefficients one should use the random effects estimator for small values of T and be indifferent between the random effects estimator and the bias-corrected fixed effects estimator for higher values of T. When marginal effects are the quantity of interest, one should use either the fixed effects estimator or the bias-corrected fixed effects estimator.

1 Introduction

Given panel data with a binary dependent variable, one usually wants to draw inference about coefficients of the latent variables and partial effects with the latter being more important in empirical applications. There is a variety of models that serve this purpose. For instance, one can apply one of the well-known models for linear panel data such as linear fixed effects, or linear random effects. The estimates correspond then to average partial effects (APE) of the model. As pointed out in Wooldridge (2010), the linear probability model (LPM) cannot be a good description of the population probability if the domain of the regressors is not severely restricted. Furthermore, Horrace and Oaxaca (2006) provide a formal proof that the ordinary least squares (OLS) estimator is biased for the LPM. Although the result corresponds to cross-sectional data, one should not expect the situation to be any different in a two-dimensional setting. Those rationales do not discourage some researchers from using OLS in datasets with binary outcomes. For instance, Klaassen and Magnus (2001) used a linear model in panel data with binary outcomes arguing their choice primarily by the restricted domain of their independent variable. The rationales for applying linear models include computational advantage and ease of interpretation.

One can also apply one of the non-linear models. Inspired by the linear fixed effects model, where one can difference out the time-invariant effects, there is a nonlinear fixed effects model (FE) based on the maximum likelihood approach (Honoré, 2002). Even though the model is flexible as it allows for an unrestricted relation between individual-specific effects and regressors, the estimates can be severely biased due to the incidental parameters problem. However, there exist bias corrections, with the notable mention of Fernández-Val (2009), that significantly reduce the bias.

One can also make additional assumptions about the distribution of the individual-specific effect, thus making the model very restrictive. One can then apply the nonlinear random effects estimator (RE). The dynamic random effects model usually suffers from the initial conditions problem. It arises if the stochastic process generating the data started before the time of the first observation (Heckman, 1981; Akay, 2009). Various estimators have been proposed which deal with the initial conditions problem (Heckman, 1981; Orme, 2001; Wooldridge, 2005). The estimator proposed by Wooldridge (2005) has high relevance in empirical work as it does not require special programming software (Arulampalam and Stewart, 2009).

Since the above-mentioned estimators suffer from various problems, it remains unclear under what circumstances one is superior to the rest. This motivates the research and gives the incentive to perform a Monte Carlo simulation analysing the performance of the models under various specifications of the data generating process. In particular, I will investigate static and dynamic probit models.

The paper is organised as follows. Section 2 reviews the literature by taking a closer look at problems encountered in various estimation techniques and proposed solutions. Section 3 describes the implementation of the Monte Carlo simulation. Section 4 presents the results. Section 5 concludes the paper by providing recommendations about the choice of the model.

2 Literature Review

2.1 Linear Models

Most of the discussion in the literature focuses on applying OLS to cross-sectional data with binary outcomes. However, there is no reason to reckon that those arguments are not valid in a two-dimensional setting. As mentioned in Horrace and Oaxaca (2006), there are plenty of empirical rationales for applying OLS on data with binary dependent variables. First of all, the obtained coefficients are easy to interpret as they correspond to APE and not the conditional expectation function. Furthermore, there are instances when OLS yields results similar to probit and logit (Angrist and Pischke, 2009; Lewbel, Dong and Yang, 2012; Klaassen and Magnus, 2001). Applying a linear model has computational benefits as it is widely available in basic software packages and it is performed fast which might play a role in very large panel datasets.

However, as indicated in Wooldridge (2010) and Lewbel et al. (2012), if the vector of observations is not severely restricted, OLS will lead to a biased and inconsistent estimator. The formal proof is given in Horrace and Oaxaca (2006). Here I briefly describe their findings.

Horrace and Oaxaca (2006) start the analysis by defining the probabilities over the random variable $x_i\beta \in \mathfrak{R}$, where x_i is a $1 \times k$ vector of explanatory variables on $\mathfrak{R}^k$ and β is a $k \times 1$ vector of coefficients.

$\Pr(x_i\beta > 1) = \pi$
$\Pr(x_i\beta \in [0,1]) = \gamma$
$\Pr(x_i\beta < 0) = \rho$

The key result of their analysis is the following:

"If $\gamma < 1$, then Ordinary Least Squares estimation of the Linear Probability Model is generally biased and inconsistent."

Horrace and Oaxaca (2006) further show that the knowledge of the probabilities γ, π, and ρ does not enable one to obtain an unbiased and consistent estimator using OLS. One would have to know the set $\kappa_\beta = \{i | x_i\beta \in [0,1]\}$. OLS performed on the observations in this set would result in an unbiased estimator. Since in practice, one does not possess the knowledge about κ_β and vector x is often not restricted to a very small set of values, OLS is biased and inconsistent in empirical applications.

2.2 Non-linear Models

2.2.1 Fixed Effects

Similarly to the linear panel data model, the non-linear fixed effects estimator based on the maximum likelihood approach does not require any assumption regarding the distribution of the individual-specific effects. However, the FE estimator suffers from the incidental parameters bias, which was first described in Neyman and Scott (1948). They proposed the following result:

"Maximum-likelihood estimates of the structural parameters relating to a partially consistent series of observations need not be consistent"

In the FE estimator, the term "incidental" refers to unobserved, individual-specific effects, whereas "structural" refers to parameters common to all individuals.

In the calculation of the FE estimator, the individual-specific effects are jointly estimated along with the structural parameters. Given that T is fixed and $N \to \infty$, the number of parameters to be estimated rises along with the sample size. The replacement of individual-specific effects by their sample counterparts contaminates the estimates of structural parameters (Fernández-Val, 2009). The FE estimator exhibits the bias is of order $O_{\mathrm{p}}(\frac{1}{T})$, both for the estimation of the coefficients and APE (Arellano and Bonhomme, 2011).

As a way to circumvent the incidental parameters problem, an estimator based on a sufficient statistic has been researched. Using it results in the conditional distribution of y_i not depending on individual-specific effects (Honoré, 2002). Furthermore, if the conditional distribution of y_i depends on the structural parameters of interest, one can obtain consistent estimates under appropriate regularity conditions (Honoré 2002; Andersen 1970). However, in most cases, a sufficient statistic is not available in non-linear panel data making the alternative infeasible (Honoré, 2002). Moreover, even if a sufficient statistic is available, one can obtain only the consistent estimates of the structural parameters without the possibility to estimate APE since in this approach one does not obtain estimates of individual-specific effects (Wooldridge, 2010).

Various bias corrections have been proposed that reduce the incidental parameter bias for both probit and logit models (Fernández-Val, 2009; Hahn and Newey's, 2004; Hahn, Kuersteiner, and Cho, 2004; Honoré and Kyriazidou's, 2000). Fernández-Val (2009) developed an asymptotic bias correction that exhibits better finite sample properties compared to other estimators. Blair and Breunig (2016) also found that this estimator is superior in estimating the coefficients. Interestingly, their simulation suggests that the uncorrected FE estimator should be preferred when APE is the quantity of interest. The bias correction can also be applied to dynamic models by specifying a bandwidth parameter. Fernandez-Val and Weidner (2018) recommend using a bandwidth not higher than 4. The bias of the bias-corrected estimator is of order $O_{\mathrm{p}}(\frac{1}{T^2})$ for the estimations of the coefficients and APE (Fernández-Val, 2009; Arellano and Bonhomme, 2011).

2.2.2 Random Effects

In the RE model, one makes certain assumptions about the distribution of the individual-specific effects. In static models, specifying a distribution of the individual-specific effects conditional on the regressors makes the model fully parametric and one can obtain estimates of the coefficients using a standard maximum likelihood approach (Honoré, 2002). In dynamic models, one would specify a parametric distribution of individual-specific effect conditional on the regressors and initial condition (Arellano and Bonhomme, 2011). However, if the stochastic process generating the data started before the process is observed, the initial condition is not exogenous anymore. Treating it as such will result in inconsistent estimates (Heckman, 1981).

Several solutions have been proposed that deal with the problem of initial conditions (Heckman, 1981; Orme, 2001; Wooldridge, 2005). Heckman (1981) proposed an estimator based on approximation of the conditional distribution of the initial condition. The usefulness of the estimator proposed by him has been limited by its requirement of special programming software (Arulampalam and Stewart, 2009). Hence, in empirical work, one often uses a dynamic probit estimator proposed by Wooldridge (2005), which can be implemented using standard programming software. There is still little evidence about the performance of an estimator proposed by Wooldridge in comparison to the one proposed by Heckman. Akay (2009) recommends using the Heckman method for short panels and the Wooldridge method for moderately long panels.

If the RE model is correctly specified, then one obtains a consistent estimator for both the coefficients and APE (Arellano and Bonhomme, 2011). Since the distributional assumptions imposed by the model are restrictive and they are motivated by convenience rather than structural and empirical evidence, there is a high chance that the model is misspecified in empirical applications. The sources of misspecifications include an incorrect assumption about the distribution of the individual-specific effects or an incorrect conditioning on exogenous covariates (Arellano and Bonhomme, 2011). The properties of the RE estimator under misspecification are similar to the FE estimator which suffers from incidental parameter bias (Arellano and Bonhomme, 2011). The bias in estimation of the coefficients is of order $O_{\mathrm{p}}(\frac{1}{T})$. However, the bias in the estimation of APE is of order $O_{\mathrm{p}}(1)$ meaning that the estimator is inconsistent as $T \to \infty$.

3 Methodology

In the binary panel data model with first-order dynamics, the response for the individual i at time t is generated by the following latent process:

$$y_{it}^{*} = \rho y_{it-1} + X_{it}'\beta + c_i + \epsilon_{it} \tag{1}$$

X_{it} denotes a $k \times 1$ vector of regressors, β is a $k \times 1$ vector of parameters, c_i is a scalar individual-specific effect, and ϵ_{it} denotes an error term. Since the estimator proposed by Wooldridge (2005) is only valid for a probit model, I make the following assumption about the error term in every data generating process:

$$\epsilon_{it} | y_{it-1}, \ldots, y_{i0}, \mathbf{x}_i, c_i \sim \text{i.i.d.N}(0, 1) \tag{2}$$

Thus the binary observation for an individual i at time t is generated by the following function:

$$\mathbb{P}\left(y_{it} = 1 | y_{it-1}, \ldots, y_{i0}, \mathbf{x}_i, c_i\right) = \Phi\left(\rho y_{it-1} + X_{it}' \beta_+ c_i\right) \tag{3}$$

I investigate the performance of the following estimation techniques: non-linear random effects (Wooldridge, 2005), non-linear fixed effects with bias correction (Fernández-Val, 2009), and linear fixed effects (e.g. Wooldridge, 2010, Section 10.5). The non-linear FE uses only individuals with time-varying responses, whereas other above-mentioned estimators use all individuals. It is interesting to investigate the performance of the RE estimator and the linear fixed effects estimator applied to the dataset consisting solely of individuals with time-varying outcomes. Consequently, I include those estimators in my simulation.

3.1 Average Partial Effects

Since partial effects are often of interest in empirical applications, one can use average partial effects as a way of investigating the performance of various estimators. This is complicated compared to the estimation of the coefficients because in every iteration of the simulation one obtains different vector of exogenous regressors that consequently impacts APE.

Some researchers (e.g. Hahn and Newey, 2004) report the partial effect from the individual with average characteristics. Fernández-Val (2009) calculates the partial effects of all the observations and averages across them. As he argues, the partial effect from an individual with average characteristics is generally not well-defined since the complete data separation might occur with positive probability. Furthermore, other researchers advocate this approach (Blair and Breunig, 2016; Chamberlain, 1984; Statmmann, 2016). Therefore, I calculate APE as an average of all partial effects taken over individuals.

Hahn and Newey (2004), Fernández-Val (2009), and Blair and Breunig (2016) report the estimation of APE as a ratio of estimated APE to true APE. Hence, I also report it as a ratio as it enables me to see the magnitude of the bias. Since the ratio is 1 for an unbiased estimator, I subsequently subtract it to find the bias.

For each data generated in the dynamic[1] model I calculate APE for both an exogenous regressor X and a lagged dependent variable y_{t-1}. Since X is continuous, the partial effect is defined as the derivative of the cumulative distribution function from equation 3.

$$\frac{\partial}{\partial_X}\mathbb{P}\left(y_{it}=1|y_{it-1},\ldots,y_{i0},\mathbf{x}_i,c_i\right)=\beta\phi\left(\rho y_{it-1}+X_{it}'\beta+c_i\right) \tag{4}$$

Note that y_{t-1} is binary. Hence, the partial effect is defined as the difference of the probabilities between different levels of y_{t-1}.

$$\Phi\left(\rho+X_{it}'\beta+c_i\right)-\Phi\left(X_{it}'\beta+c_i\right) \tag{5}$$

By taking the sample mean of the corresponding partial effects, I obtain APE for X and y_{t-1}.

$$APE_X=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\beta\phi\left(\rho+X_{it}'\beta+c_i\right) \tag{6}$$

$$APE_Y=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\Phi\left(\rho+X_{it}'\beta+c_i\right)-\Phi\left(X_{it}'\beta+c_i\right)\right] \tag{7}$$

[1]For the static model I calculate APE_x

3.2 Dynamic Probit Model by Wooldridge

I provide a brief description of an estimator proposed by Wooldridge (2005). Assume the probit model with first-order dynamics as shown in equation 3. Due to convenience, Wooldridge assumes the following conditional density of the individual-specific effects[2]

$$c_i|y_{i0}, \mathbf{x}_i \sim \mathrm{N}\left(\alpha_0 + \alpha_1 y_{i0} + \mathbf{x}_i\boldsymbol{\alpha}_2, \sigma_a^2\right) \tag{8}$$

Since both the error term and the individual-specific effects are independent and normally distributed random variables, their sum also follows a normal distribution. This enables Wooldridge to write the probit model as[3]:

$$\Phi\left(\rho y_{i,t-1} + X_{it}'\beta + \alpha_0 + \alpha_1 y_{i0} + \mathbf{x}_i\boldsymbol{\alpha}_2 + a_i\right) \tag{9}$$

The conditional joint density can be then written as:

$$\begin{aligned}\prod_{t=1}^{T} \Big\{ &\Phi\left(\rho y_{i,t-1} + X_{it}'\beta + \alpha_0 + \alpha_1 y_{i0} + \mathbf{x}_i\boldsymbol{\alpha}_2 + a_i\right)^{y_t} \\ &\times \left[1 - \Phi\left(\rho y_{i,t-1} + X_{it}'\beta + \alpha_0 + \alpha_1 y_{i0} + \mathbf{x}_i\boldsymbol{\alpha}_2 + a_i\right)\right]^{1-y_t} \Big\}\end{aligned} \tag{10}$$

Integrating out the equation against the conditional distribution of α_i gives:

$$\begin{aligned}&\int_{\mathbf{R}} \Big(\prod_{t=1}^{T} \Big\{ \Phi\left(\rho y_{i,t-1} + X_{it}'\beta + \alpha_0 + \alpha_1 y_{i0} + \mathbf{x}_i\boldsymbol{\alpha}_2 + a_i\right)^{y_t} \\ &\times \left[1 - \Phi\left(\rho y_{i,t-1} + X_{it}'\beta + \alpha_0 + \alpha_1 y_{i0} + \mathbf{x}_i\boldsymbol{\alpha}_2 + a_i\right)\right]^{1-y_t} \Big) \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha}{\sigma_\alpha}\right) \mathrm{d}\alpha\end{aligned} \tag{11}$$

Taking the logarithm of the maximum likelihood function, one obtains the log-likelihood function which can be optimized numerically. One can then obtain the maximum likelihood estimators of $\hat{\beta}, \hat{\rho}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\boldsymbol{\alpha}}_2$ and $\hat{\sigma}_a$.

[2]One cane also write it as $c_i = \alpha_0 + \alpha_1 y_{i0} + \mathbf{x}_i\boldsymbol{\alpha}_2 + \alpha_i$ with $\alpha_i|y_{i0}, \mathbf{x}_i \sim \mathrm{N}\left(0, \sigma_a^2\right)$

[3]To see it, note that one can write the latent process as $y_{it}^* = \rho y_{i,t-1} + X_{it}'\beta + c_i + \epsilon_{it}$

Wooldridge allows for the individual-specific effects to be dependent on the whole vector $\mathbf{x}_i$. In empirical applications, one often assumes that the individual effect is dependent on the average of the observations. Consequently, I apply the dynamic random effects probit model with the following assumption about the individual-specific effects:

$$c_i|y_{i0}, \bar{X}_i \sim \mathrm{N}\left(\alpha_0 + \alpha_1 y_{i0} + \bar{X}_i\alpha_2, \sigma_a^2\right) \tag{12}$$

Having obtained the estimated coefficients, the estimations of APE of X and ${y_{t-1}}^4$ are given by:

$$\begin{aligned} A\hat{P}E_Y = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Phi\left(\hat{\rho}_a + X_{it}'\hat{\beta}_a + \hat{\alpha}_{a0} + \hat{\alpha}_{a1}y_{i0} + \bar{X}_i'\hat{\alpha}_{a3}\right) \\ -\Phi\left(X_{it}'\hat{\beta}_a + \hat{\alpha}_{a0} + \hat{\alpha}_{a1}y_{i0} + \bar{X}_i'\hat{\alpha}_{a3}\right) \end{aligned} \tag{13}$$

$$A\hat{P}E_X = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\beta}_a\phi\left(\hat{\rho}_a + X_{it}'\hat{\beta}_a + \hat{\alpha}_{a0} + \hat{\alpha}_{a1}y_{i0} + \bar{X}_i'\hat{\alpha}_{a3}\right) \tag{14}$$

Note that the static RE probit model does not suffer from the initial conditions problem. Hence, for data without dynamics I apply the random effects estimator with the following assumption about the distribution of the individual-specific effects:

$$c_i|\bar{X}_i \sim \mathrm{N}\left(\alpha_0 + \bar{X}_i\alpha_2, \sigma_a^2\right) \tag{15}$$

3.3 FE

The FE estimator in the simulation corresponds to the unconditional probit estimator. In order to fit the model efficiently, I used the pseudo demeaning algorithm derived by Stammann, Heiss, and McFadden (2016). Having obtained the estimated coefficients and individual-specific effects, I obtain the

[4] Subscript α indicates the multiplication of the estimated coefficient by $1/\sqrt{(1+\hat{\sigma}_\alpha^2}$

estimated APE.

$$A\hat{P}E_Y = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Phi\left(\hat{\rho} + X'_{it}\hat{\beta} + \hat{\alpha}_i(\beta)\right) - \Phi\left(X'_{it}\hat{\beta} + \hat{\alpha}_i(\beta)\right) \tag{16}$$

$$A\hat{P}E_X = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\beta}\phi\left(\hat{\rho} + X'_{it}\hat{\beta} + \hat{\alpha}_i(\beta)\right) \tag{17}$$

I apply the post-estimation routine of Fernández-Val (2009) to reduce the incidental parameter bias and subsequently calculate APE with the newly estimated coefficients. In dynamic models, I also investigate the performance of bias correction for the various values of bandwidth parameters.

The simulation was performed in R using 'plm', 'pglm', and 'bife' packages. The numerical maximisations were accomplished using the Newton-Raphson algorithm. The implementation of APE for the RE estimator was unsuccessful as 'pglm' does not estimate the standard deviation of the individual effect, σ_α, correctly. I briefly describe the problem in the appendix. Consequently, I do not present the results for APE of the RE estimator since they are invalid.

4 Results

Throughout the results, MSE stands for mean squared error, MAE for mean absolute error, SD for the standard deviation of an estimator, $p;x$ is the rejection frequency with x being the nominal value. The rejection frequencies for all the estimators were obtained using classical statistical testing[5]. RE denotes the random effects estimator (Wooldridge, 2005), RE(R) denotes the random effects estimator performed on time-varying individuals, FE denotes the fixed effects estimator, FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009) with the number in the bracket indicating

[5] I implemented bootstrap for APE of the RE estimator but since the estimations are invalid I do not report it either

the bandwidth parameter, LPM denotes the linear fixed effect estimator, LPM(R) denotes the linear fixed effect estimator performed on time-varying individuals.

In this section, I report the results for static and dynamic models where the RE estimator is misspecified. The misspecification is a result of both the incorrect assumption about the distribution of the individual-specific effects and the incorrect conditioning on exogenous variables. The results for the static and dynamic models where the RE estimator is correctly specified are in the appendix. To ensure the validity of the results, I replicated table 3 from Fernández-Val (2009). The replicated results are also located in the appendix.

4.1 Misspecified Static Probit Model

For the static model, I use the Nerlove process as it enables me to compare the results to previous studies. The data generating process (DGP) is characterised below. Notice that the difference between DGP used in this paper and the one used by Fernández-Val (2009) is the correlation between the exogenous variable and the individual-specific effect, and the distribution of the random component of the individual-specific effect.

$$
\begin{aligned}
&y_{\mathrm{it}} = 1\left\{X_{\mathrm{it}}\beta + c_i + \epsilon_{\mathrm{it}} \geq 0\right\} \\
&X_{it} = t/10 + X_{i,t-1}/2 + u_{it} \text{ for } t = 1, \ldots, T; \\
&X_{i0} = u_{i0}; \quad u_{it} \sim i.i.d.U(-1/2, 1/2) \\
&\epsilon_{\mathrm{it}} \sim i.i.d.N(0,1); \\
&c_{\mathrm{i}} = \alpha_1 x_{i0} + \alpha_{\mathrm{i}} \quad \alpha_{\mathrm{i}} \sim i.i.d.Exp(1) \\
&n = 250; \quad T = 4, 8, 12 \\
&\beta = 1; \quad \alpha_1 = 1
\end{aligned}
\tag{18}
$$

Table 1: Misspecified Static Probit; β

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
RE	0.06	0.06	0.05	0.18	0.21	0.00	0.02
RE(R)	0.27	0.27	0.13	0.30	0.25	0.04	0.15
FE	0.46	0.46	0.30	0.47	0.30	0.49	0.60
FE(BC)	0.09	0.09	0.05	0.18	0.21	0.05	0.08
T=8							
RE	0.04	0.04	0.01	0.09	0.10	0.02	0.07
RE(R)	0.08	0.08	0.02	0.11	0.10	0.04	0.15
FE	0.23	0.23	0.07	0.23	0.13	0.53	0.65
FE(BC)	0.04	0.04	0.01	0.09	0.10	0.04	0.09
T=12							
RE	0.03	0.04	0.01	0.07	0.08	0.04	0.09
RE(R)	0.05	0.05	0.01	0.07	0.08	0.07	0.13
FE	0.19	0.19	0.04	0.19	0.10	0.59	0.69
FE(BC)	0.03	0.03	0.01	0.07	0.08	0.05	0.10

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009).

Table 2: Misspecified Static Probit; APE_x

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
FE	-0.01	0.00	0.03	0.14	0.18	0.00	0.00
FE(BC)	0.16	0.17	0.06	0.21	0.19	0.01	0.03
LPM	0.11	0.11	0.05	0.19	0.20	0.11	0.20
LPM(R)	1.36	1.37	2.03	1.36	0.42	0.89	0.95
T=8							
FE	-0.03	-0.03	0.01	0.07	0.09	0.00	0.00
FE(BC)	0.07	0.07	0.01	0.09	0.09	0.00	0.01
LPM	0.08	0.08	0.02	0.10	0.10	0.16	0.26
LPM(R)	1.05	1.05	1.13	1.05	0.18	1.00	1.00
T=12							
FE	-0.03	-0.03	0.01	0.06	0.07	0.00	0.00
FE(BC)	0.05	0.05	0.01	0.07	0.07	0.00	0.00
LPM	0.01	0.01	0.00	0.06	0.07	0.11	0.17
LPM(R)	0.95	0.95	0.92	0.95	0.13	1.00	1.00

500 replications. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009). LPM denotes the linear fixed effect estimator. LPM(R) denotes the linear fixed effect estimator performed on time-varying individuals.

4.1.1 Estimations of Coefficients

U and Exp denote uniform and exponential distributions respectively. Table 1 presents the results for the estimators of the coefficient β. As in the simulation performed by Fernández-Val (2009), the bias of FE is severe, even for $T = 12$. Note that the bias of FE also contaminates the rejection frequencies. Same as in the simulations performed by Blair and Breunig (2016) and Fernández-Val (2009), FE(BC) performs well in reducing the bias. One can see that RE and FE(BC) exhibit similar results. In particular, MSE, MAE, and SD are very close for those two estimators for all values of T. For smaller T, FE(BC) has a higher mean bias. However, this is compensated by better results in rejection frequencies. For other distributions of the individual-specific effects[6], RE and FE(BC) also exhibit similar properties in estimating the coefficient β.

RE(R) performs much worse than RE for smaller values of T. This is driven by the data loss, as many individuals do not change the status over time if T is small. For high values of T, RE(R) performs similarly to RE, which is not surprising since data loss is not severe. However, RE is superior to RE(R) as it has a lower mean bias, and rejections frequencies are closer to the nominal value.

4.1.2 Estimations of Average Partial Effects

Table 2 presents the results for the estimators of APE_x. FE exhibits very small bias for $T = 4$, and FE(BC) increases the bias. The same result can be observed in the correctly specified model in the appendix. This is consistent with the simulations on the static probit model performed by Blair and Breunig (2016) and Fernández-Val (2009). It raises a serious concern about the performance of the bias correction in the estimation of APE. One can also observe a poor performance of FE and FE(BC) in terms of rejection

[6] In particular: T-distribution (DF=2), T-distribution (DF=100), Poisson($\lambda = 1$), Uniform(-2,2)

frequencies. Agreeing with the results of Fernández-Val (2009), LPM does fairly well in estimating APE_x in terms of mean bias. It performs better than FE and FE(BC) in rejection frequencies. In other static models investigated[7], LPM in general exhibits small bias. Note that LPM(R) performs poorly in all aspects.

4.2 Misspecified Dynamic Probit Model

Neither Fernández-Val (2009) nor Blair and Breunig (2016) ran experiments on a dynamic probit model. Hence, no direct comparison to other studies can be made. To investigate a different setting, I eliminate dynamics from the exogenous variable and change the distribution of the individual-specific effects. Notice that the misspecification of RE is severe as the distribution of c_i is far from the one assumed in the estimator.

$$\begin{aligned} & y_{\mathrm{it}} = 1\left\{\rho y_{it-1} + X_{\mathrm{it}}\beta + c_i + \epsilon_{\mathrm{it}} \geq 0\right\} \\ & X_{\mathrm{it}} \sim i.i.d.N(0,1) \\ & \epsilon_{\mathrm{it}} \sim i.i.d.N(0,1); \\ & c_{\mathrm{i}} = \alpha_0 + \alpha_1 x_{i0} + \alpha_{\mathrm{i}} \quad \alpha_{\mathrm{i}} \sim i.i.d.U(-2,2) \\ & n = 250; \quad T = 4, 8, 12 \\ & \rho = 1; \quad \beta = 1; \quad \alpha_0 = -1; \quad \alpha_1 = 1 \end{aligned} \tag{19}$$

4.2.1 Estimations of Coefficients

Tables 3 and 4 present the results for the estimators of ρ and β respectively. Note that there is a large discrepancy in the performance of the bias correction for various values of the bandwidth parameter when T is small. One can observe that for $T = 4$, a unit bandwidth parameter is superior. For larger T, the discrepancy becomes smaller with FE(BC) exhibiting the best properties when the bandwidth parameter is set to either 1 or 2. This result holds

[7] In particular: T-distribution (DF=2), T-distribution (DF=100), Poisson($\lambda = 1$), Uniform(-2,2)

Table 3: Misspecified Dynamic Probit; ρ

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
RE	0.14	0.14	0.06	0.19	0.19	0.01	0.08
RE(R)	-0.34	-0.34	0.14	0.34	0.15	0.55	0.65
FE	-0.73	-0.73	0.60	0.73	0.25	0.56	0.81
FE(BC-0)	-0.85	-0.85	0.75	0.85	0.14	0.96	1.00
FE(BC-1)	-0.27	-0.27	0.09	0.27	0.15	0.03	0.14
FE(BC-2)	-0.54	-0.54	0.32	0.54	0.17	0.40	0.73
FE(BC-3)	-1.08	-1.08	1.19	1.08	0.14	1.00	1.00
T=8							
RE	0.05	0.04	0.01	0.09	0.11	0.03	0.08
RE(R)	-0.04	-0.04	0.01	0.09	0.10	0.04	0.10
FE	-0.33	-0.34	0.12	0.33	0.12	0.66	0.80
FE(BC-0)	-0.45	-0.45	0.22	0.45	0.10	0.97	0.99
FE(BC-1)	-0.10	-0.10	0.02	0.11	0.10	0.04	0.13
FE(BC-2)	-0.09	-0.09	0.02	0.12	0.11	0.07	0.15
FE(BC-3)	-0.16	-0.15	0.04	0.17	0.12	0.17	0.30
T=12							
RE	0.02	0.02	0.01	0.07	0.09	0.03	0.09
RE(R)	-0.01	-0.01	0.01	0.07	0.09	0.05	0.10
FE	-0.22	-0.22	0.06	0.22	0.10	0.58	0.72
FE(BC-0)	-0.32	-0.32	0.11	0.32	0.08	0.92	0.97
FE(BC-1)	-0.07	-0.07	0.01	0.09	0.08	0.06	0.16
FE(BC-2)	-0.05	-0.04	0.01	0.08	0.09	0.04	0.10
FE(BC-3)	-0.07	-0.06	0.01	0.09	0.09	0.07	0.15

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009).

Table 4: Misspecified Dynamic Probit; β

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
RE	-0.01	-0.02	0.01	0.08	0.11	0.00	0.03
RE(R)	0.14	0.13	0.03	0.14	0.11	0.19	0.30
FE	0.56	0.54	0.36	0.56	0.22	0.61	0.86
FE(BC-0)	-0.23	-0.22	0.09	0.25	0.19	0.20	0.43
FE(BC-1)	-0.18	-0.18	0.05	0.19	0.14	0.11	0.26
FE(BC-2)	-0.24	-0.24	0.10	0.27	0.20	0.27	0.53
FE(BC-3)	-0.29	-0.29	0.12	0.31	0.20	0.42	0.72
T=8							
RE	-0.00	-0.01	0.00	0.05	0.06	0.01	0.06
RE(R)	0.02	0.02	0.00	0.05	0.07	0.05	0.09
FE	0.24	0.23	0.07	0.24	0.09	0.75	0.88
FE(BC-0)	0.00	-0.00	0.00	0.05	0.06	0.00	0.02
FE(BC-1)	0.02	0.02	0.00	0.05	0.06	0.01	0.05
FE(BC-2)	0.02	0.01	0.00	0.05	0.06	0.00	0.03
FE(BC-3)	0.01	0.01	0.00	0.05	0.06	0.00	0.02
T=12							
RE	0.00	-0.00	0.00	0.04	0.05	0.03	0.08
RE(R)	0.02	0.02	0.00	0.04	0.05	0.06	0.10
FE	0.16	0.16	0.03	0.16	0.07	0.71	0.83
FE(BC-0)	0.01	0.01	0.00	0.04	0.05	0.02	0.06
FE(BC-1)	0.02	0.02	0.00	0.04	0.05	0.04	0.08
FE(BC-2)	0.02	0.02	0.00	0.05	0.05	0.03	0.07
FE(BC-3)	0.02	0.02	0.00	0.05	0.05	0.02	0.08

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009).

for estimations of both coefficients. The same result can be observed in the dynamic probit model where RE is correctly specified.

For $T = 4$, RE is superior in the estimations of both ρ and β. Similarly, as in the static model, the bias of FE is severe and the bias correction significantly reduces it. There is a large difference in the performance of FE(BC) if one increases T from 4 to 8. This is not a surprise considering that FE(BC) exhibits a bias of order $O_{\mathrm{p}}(\frac{1}{T^2})$. Note that RE and FE(BC) exhibit similar properties for $T = 8$ and $T = 12$. This result was also observed for other distributions of individual effect[8].

Similarly, as in the static model, RE(R) performs poorly for small T, and for larger T it exhibits similar properties to RE. Interestingly, RE(R) performs better in terms of rejection frequencies for larger T. This result can be also observed in the dynamic probit model where RE is correctly specified.

4.2.2 Estimations of Average Partial Effects

Tables 5 and 6 present the results of the estimators of APE_y and APE_x respectively. Similarly, as in the estimation of the coefficients, there are large discrepancies in the performance of bias-corrected FE for different values of the bandwidth parameter. Once more, a unit bandwidth is superior for $T = 4$, whereas, for larger values of T, FE(BC-2) exhibits slightly better properties than other estimators.

In table 5, one can see that with the properly chosen bandwidth parameter, FE(BC) is superior for all values of T as an estimator of APE_y. Note that LPM exhibits very large bias for small T. This also has an impact on the rejection frequencies. A similar result can be observed in the simulation performed on the dynamic logit model by Fernández-Val (2009).

LPM does significantly better in estimating APE_x compared to APE_y. For

[8] In particular: Exponential($\lambda = 1$), T-distribution (DF=2), T-distribution (DF=100) Poisson($\lambda = 1$), Discrete Uniform[-2,2]

Table 5: Misspecified Dynamic Probit; APE_y

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
FE	-0.88	-0.88	0.78	0.88	0.12	0.74	0.99
FE(BC-0)	-0.86	-0.86	0.76	0.86	0.13	0.61	0.96
FE(BC-1)	-0.33	-0.34	0.13	0.33	0.13	0.01	0.03
FE(BC-2)	-0.58	-0.57	0.36	0.58	0.15	0.06	0.37
FE(BC-3)	-1.07	-1.07	1.16	1.07	0.12	0.95	1.00
LPM	-0.90	-0.91	0.84	0.90	0.16	1.00	1.00
LPM(R)	-0.77	-0.78	0.65	0.77	0.26	0.76	0.87
T=8							
FE	-0.54	-0.54	0.30	0.54	0.08	0.99	1.00
FE(BC-0)	-0.52	-0.51	0.28	0.52	0.10	0.95	0.99
FE(BC-1)	-0.19	-0.19	0.04	0.19	0.09	0.06	0.20
FE(BC-2)	-0.18	-0.17	0.05	0.19	0.11	0.04	0.16
FE(BC-3)	-0.25	-0.24	0.07	0.25	0.11	0.14	0.33
LPM	-0.37	-0.38	0.15	0.37	0.12	0.91	0.95
LPM(R)	-0.26	-0.27	0.09	0.27	0.14	0.43	0.59
T=12							
FE	-0.40	-0.40	0.17	0.40	0.08	0.98	1.00
FE(BC-0)	-0.39	-0.39	0.16	0.39	0.08	0.93	0.98
FE(BC-1)	-0.13	-0.12	0.02	0.13	0.08	0.11	0.22
FE(BC-2)	-0.11	-0.11	0.02	0.12	0.09	0.04	0.14
FE(BC-3)	-0.13	-0.13	0.03	0.14	0.09	0.07	0.18
LPM	-0.18	-0.18	0.04	0.18	0.11	0.51	0.61
LPM(R)	-0.11	-0.11	0.02	0.13	0.11	0.18	0.27

500 replications. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009) with the number in the bracket indicating the bandwidth parameter. LPM denotes the linear fixed effect estimator. LPM(R) denotes the linear fixed effect estimator performed on time-varying individuals.

Table 6: Misspecified Dynamic Probit; APE_x

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
FE	-0.30	-0.30	0.09	0.30	0.06	0.00	0.17
FE(BC-0)	-0.29	-0.29	0.09	0.29	0.07	0.00	0.06
FE(BC-1)	-0.26	-0.25	0.07	0.26	0.09	0.07	0.31
FE(BC-2)	-0.33	-0.32	0.12	0.33	0.10	0.16	0.53
FE(BC-3)	-0.40	-0.39	0.17	0.40	0.10	0.38	0.84
LPM	-0.26	-0.26	0.07	0.26	0.07	0.96	0.99
LPM(R)	0.67	0.67	0.46	0.67	0.11	1.00	1.00
T=8							
FE	-0.22	-0.22	0.05	0.22	0.05	0.81	0.96
FE(BC-0)	-0.19	-0.19	0.04	0.19	0.04	0.55	0.84
FE(BC-1)	-0.17	-0.17	0.03	0.17	0.04	0.57	0.81
FE(BC-2)	-0.17	-0.17	0.03	0.17	0.04	0.27	0.58
FE(BC-3)	-0.18	-0.18	0.03	0.18	0.04	0.38	0.69
LPM	-0.21	-0.21	0.05	0.21	0.05	1.00	1.00
LPM(R)	0.29	0.29	0.09	0.29	0.06	1.00	1.00
T=12							
FE	-0.21	-0.21	0.04	0.21	0.04	0.98	0.99
FE(BC-0)	-0.19	-0.19	0.04	0.19	0.04	0.92	0.97
FE(BC-1)	-0.18	-0.18	0.03	0.18	0.04	0.95	0.98
FE(BC-2)	-0.17	-0.17	0.03	0.17	0.04	0.80	0.94
FE(BC-3)	-0.18	-0.18	0.03	0.18	0.04	0.83	0.95
LPM	-0.20	-0.20	0.04	0.20	0.04	1.00	1.00
LPM(R)	0.16	0.16	0.03	0.16	0.05	0.93	0.96

500 replications. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009) with the number in the bracket indicating the bandwidth parameter. LPM denotes the linear fixed effect estimator. LPM(R) denotes the linear fixed effect estimator performed on time-varying individuals.

$T = 4$, the performance of LPM is similar to FE(BC-1) except for rejection frequencies. However, for larger values of T, FE(BC) is superior to LPM. Surprisingly, for larger values of T, FE(BC) performs worse in terms of rejection frequencies than for small values T.

Notice that in general, FE(BC) exhibits worse properties than in the simulations performed by Fernández-Val (2009). This is likely a result of the more complicated data generating process used in this paper. However, considering that FE(BC) did not perform well in reducing the bias of APE in the reported static model and similar instances were found by Blair and Breunig (2016), it raises a serious concern about the performance of bias correction in the estimation of APE in static models. Nevertheless, the bias correction performed well in estimations of APE in dynamic probit models. Furthermore, agreeing with Blair and Breunig (2016), I did not find an instance where FE outperformed FE(BC) in estimating the coefficients.

5 Conclusion

I have performed a Monte-Carlo simulation investigating the performance of the linear probability model, non-linear fixed effects model with bias correction proposed by Fernández-Val (2009), and non-linear random effects model proposed by Wooldridge (2005). The results indicate that when applying the bias correction for the FE estimator in the dynamic probit model, one should use a unit bandwidth parameter for small values of T, whereas a bandwidth parameter of 1 or 2 for higher values of T is preferable. The results suggest that when estimating the coefficients, one should apply RE for small values of T and should be indifferent between RE and bias-corrected FE for higher values of T. This is explained by the severity of the bias for small values of T but a higher rate of convergence of FE(BC) compared to RE.

The simulation shows that LPM works well in a simple setting but is severely biased in more complicated models. Consequently, for the estimation of APE, one should apply either FE or bias-corrected FE. Since uncorrected FE performed better static models and bias-corrected FE performed better in dynamic models in the estimation of APE, it remains unclear which model is preferred. This gives an incentive for further research about the behaviour of the bias correction in the estimation of APE. Another interesting issue is the performance of RE carried out on individuals with time-varying outcomes for higher values of T. The simulation suggests that in some cases this estimator can perform better in rejection frequencies compared to RE performed on the whole dataset.

One severe limitation of the paper is the lack of implementation of APE for the RE model. Since the estimator of APE, when RE is misspecified, has a bias of $O_P(1)$, it already indicates that one should prefer either FE or bias-corrected FE for higher values of T. However, it remains unclear which estimator should be used for small values of T.

References

[1] Alpaslan Akay. The wooldridge method for the initial values problem is simple: what about performance? 2009.

[2] Blair Alexander and Robert Breunig. A monte carlo study of bias corrections for panel probit models. *Journal of Statistical Computation and Simulation*, 86(1):74–90, 2016.

[3] Erling Bernhard Andersen. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(2):283–301, 1970.

[4] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

[5] Manuel Arellano and Stéphane Bonhomme. Nonlinear panel data analysis. 2011.

[6] Wiji Arulampalam and Mark B Stewart. Simplified implementation of the heckman estimator of the dynamic probit model and a comparison with alternative estimators. *Oxford bulletin of economics and statistics*, 71(5):659–681, 2009.

[7] Gary Chamberlain. Panel data. *Handbook of econometrics*, 2:1247–1318, 1984.

[8] Iván Fernández-Val. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85, 2009.

[9] Iván Fernández-Val and Martin Weidner. Fixed effects estimation of large-t panel data models. *Annual Review of Economics*, 10:109–138, 2018.

[10] Jinyong Hahn, Guido Kuersteiner, and Myeong Hyeon Cho. Asymptotic distribution of misspecified random effects estimator for a dynamic panel model with fixed effects when both n and t are large. *Economics Letters*, 84(1):117–125, 2004.

[11] Jinyong Hahn and Whitney Newey. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319, 2004.

[12] James Heckman. J. 1981, the incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process.

[13] Bo E Honoré. Nonlinear models with panel data. *Portuguese Economic Journal*, 1(2):163–179, 2002.

[14] Bo E Honoré and Ekaterini Kyriazidou. Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874, 2000.

[15] William C Horrace and Ronald L Oaxaca. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90(3):321–327, 2006.

[16] Franc JGM Klaassen and Jan R Magnus. Are points in tennis independent and identically distributed? evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96(454):500–509, 2001.

[17] Arthur Lewbel, Yingying Dong, and Thomas Tao Yang. Comparing features of convenient estimators for binary choice models with endogenous regressors. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(3):809–829, 2012.

[18] Jerzy Neyman and Elizabeth L Scott. Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32, 1948.

[19] Chris D Orme. Two-step inference in dynamic non-linear panel data models. *Unpublished paper, University of Manchester. http://personalpages. manchester. ac. uk/staff/chris. orme/documents/Research% 20Paper s/initcondlast. pdf*, 2001.

[20] Amrei Stammann, Florian Heiss, and Daniel McFadden. Estimating fixed effects logit models with large panel data. 2016.

[21] Jeffrey M Wooldridge. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics*, 20(1):39–54, 2005.

[22] Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data.* 2010.

6 Appendix

6.1 The Source of Bias in APE of RE

Wooldridge (2005) states that one needs to obtain MLEs of all the coefficients and σ_α and then subsequently use them to calculate APE. Here I show that the function 'pglm' in R does not return the estimates of σ_α which in turn invalidates APE of RE. I simulate the data from a simple, static probit model:

$$
\begin{aligned}
& y_{\text{it}} = 1\left\{X_{\text{it}}\beta + c_i + \epsilon_{\text{it}} \geq 0\right\} \\
& \epsilon_{\text{it}} \sim i.i.d.N(0,1); \\
& c_{\text{i}} = \alpha_0 + \alpha_{\text{i}} \sim i.i.d.N(0,1) \\
& X_{\text{it}} \sim i.i.d.N(0,1) \\
& n = 250; \quad T = 8; \quad \beta = 1; \quad \alpha_0 = 1
\end{aligned} \tag{20}
$$

Note that since $\alpha_{\text{i}} \sim i.i.d.N(0, \sigma_a)$, random effects estimator is correctly specified and 'pglm' should report unbiased estimates of all parameters. Table 7 presents average estimates of coefficients from 'pglm' package. One can observe that all the coefficients are estimated correctly except for σ_α. Note that the problem does not disappear if I increase the number of replications, individuals, or time.

6.2 Correctly Specified Static Probit Model

$$
\begin{aligned}
& y_{\text{it}} = 1\left\{X_{\text{it}}\beta + c_i + \epsilon_{\text{it}} \geq 0\right\} \\
& X_{it} = t/10 + X_{i,t-1}/2 + u_{it} \text{ for } t = 1, \ldots, T; \\
& X_{i0} = u_{i0}; \quad u_{it} \sim i.i.d.U(-1/2, 1/2) \\
& \epsilon_{\text{it}} \sim i.i.d.N(0,1); \\
& c_{\text{i}} = \alpha_0 + \alpha_1 \bar{x}_i + \alpha_{\text{i}} \quad \alpha_{\text{i}} \sim i.i.d.N(0,1) \\
& n = 250; \quad T = 4, 8, 12 \\
& \beta = 1; \quad \alpha_0 = -1; \quad \alpha_1 = 1
\end{aligned} \tag{21}
$$

Table 7: Bias in Estimation of σ_a

	Intercept	β	σ_a
$\sigma_a = 0.1$	1.01	1.004	-0.143
$\sigma_a = 0.3$	1.002	0.997	-0.027
$\sigma_a = 0.5$	0.996	0.998	0.674
$\sigma_a = 1$	0.995	0.995	1.401
$\sigma_a = 2$	0.974	0.989	2.767

100 replications

Table 8: Correctly Specified Static Probit; β

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
RE	0.01	0.02	0.03	0.13	0.16	0.00	0.01
RE(R)	0.18	0.18	0.07	0.21	0.18	0.01	0.07
FE	0.42	0.42	0.23	0.43	0.24	0.54	0.65
FE(BC)	0.07	0.07	0.03	0.14	0.17	0.03	0.07
T=8							
RE	-0.00	-0.00	0.01	0.06	0.07	0.02	0.06
RE(R)	0.01	0.01	0.01	0.06	0.07	0.01	0.05
FE	0.17	0.17	0.04	0.17	0.09	0.54	0.68
FE(BC)	0.01	0.01	0.01	0.06	0.07	0.03	0.09
T=12							
RE	0.00	0.00	0.00	0.04	0.05	0.03	0.06
RE(R)	0.01	0.01	0.00	0.04	0.05	0.03	0.06
FE	0.13	0.13	0.02	0.13	0.06	0.59	0.74
FE(BC)	0.01	0.01	0.00	0.04	0.05	0.05	0.09

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009).

Table 9: Correctly Specified Static Probit; APE_x

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
FE	0.01	0.01	0.02	0.12	0.15	0.00	0.02
FE(BC)	0.07	0.08	0.03	0.13	0.15	0.02	0.04
LPM	0.03	0.04	0.02	0.13	0.15	0.05	0.09
LPM(R)	0.72	0.72	0.58	0.72	0.25	0.77	0.86
T=8							
FE	-0.01	-0.01	0.00	0.05	0.06	0.01	0.04
FE(BC)	0.00	0.00	0.00	0.05	0.06	0.02	0.04
LPM	0.01	0.01	0.00	0.05	0.07	0.05	0.09
LPM(R)	0.28	0.28	0.08	0.28	0.08	0.91	0.95
T=12							
FE	-0.02	-0.01	0.00	0.03	0.04	0.00	0.01
FE(BC)	0.00	0.01	0.00	0.03	0.04	0.00	0.02
LPM	0.00	-0.00	0.00	0.04	0.05	0.05	0.09
LPM(R)	0.35	0.34	0.12	0.35	0.06	1.00	1.00

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009).

Table 10: Replicating Table 3 from Fernández-Val (2009)

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
RE	0.01	-0.00	0.07	0.21	0.27	0.00	0.02
RE(R)	0.17	0.16	0.12	0.27	0.30	0.00	0.04
FE	0.41	0.39	0.33	0.46	0.39	0.27	0.36
FE(BC)	0.06	0.05	0.08	0.22	0.28	0.03	0.07
T=8							
RE	0.00	-0.01	0.02	0.10	0.13	0.02	0.07
RE(R)	0.02	0.01	0.02	0.10	0.13	0.02	0.06
FE	0.18	0.17	0.06	0.19	0.16	0.26	0.37
FE(BC)	0.01	0.01	0.02	0.10	0.13	0.05	0.10
T=12							
RE	0.00	-0.00	0.01	0.07	0.08	0.02	0.06
RE(R)	0.01	0.01	0.01	0.06	0.08	0.02	0.06
FE	0.13	0.12	0.03	0.14	0.10	0.28	0.40
FE(BC)	0.01	0.01	0.01	0.07	0.08	0.04	0.09

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009).

6.3 Correctly Specified Dynamic Probit

$$\begin{aligned}
&y_{\mathrm{it}} = 1\left\{\rho y_{it-1} + X_{\mathrm{it}}\beta + c_i + \epsilon_{\mathrm{it}} \geq 0\right\} \\
&X_{\mathrm{it}} \sim i.i.d.N(0,1) \text{ for } t = 1, \ldots, T \\
&\epsilon_{\mathrm{it}} \sim i.i.d.N(0,1) \\
&c_{\mathrm{i}} = \alpha_0 + \alpha_1 \bar{x}_i + \alpha_2 y_{i0} + \alpha_{\mathrm{i}} \quad \alpha_{\mathrm{i}} \sim i.i.d.N(0,1) \\
&n = 250; \quad T = 4, 8, 12 \\
&\rho = 1; \quad \beta = 1; \quad \alpha_0 = -1; \quad \alpha_1 = 1; \quad \alpha_2 = 1
\end{aligned} \tag{22}$$

Table 11: Correctly Specified Dynamic Probit; ρ

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
RE	-0.00	-0.00	0.03	0.14	0.18	0.01	0.03
RE(R)	-0.37	-0.38	0.16	0.37	0.15	0.69	0.80
FE	-0.75	-0.76	0.63	0.76	0.25	0.67	0.87
FE(BC-0)	-0.88	-0.88	0.80	0.88	0.13	1.00	1.00
FE(BC-1)	-0.29	-0.29	0.10	0.29	0.14	0.05	0.24
FE(BC-2)	-0.58	-0.58	0.36	0.58	0.15	0.64	0.89
FE(BC-3)	-1.11	-1.11	1.25	1.11	0.13	1.00	1.00
T=8							
RE	0.00	0.00	0.01	0.08	0.10	0.02	0.07
RE(R)	-0.05	-0.05	0.01	0.09	0.10	0.08	0.14
FE	-0.34	-0.35	0.13	0.34	0.11	0.77	0.87
FE(BC-0)	-0.46	-0.47	0.22	0.46	0.09	0.99	1.00
FE(BC-1)	-0.11	-0.11	0.02	0.12	0.09	0.07	0.19
FE(BC-2)	-0.10	-0.10	0.02	0.12	0.10	0.08	0.18
FE(BC-3)	-0.17	-0.17	0.04	0.17	0.10	0.24	0.42
T=12							
RE	-0.00	-0.00	0.01	0.06	0.08	0.02	0.07
RE(R)	-0.02	-0.02	0.01	0.06	0.07	0.04	0.11
FE	-0.23	-0.23	0.06	0.23	0.08	0.74	0.84
FE(BC-0)	-0.32	-0.32	0.11	0.32	0.07	0.97	0.99
FE(BC-1)	-0.07	-0.08	0.01	0.09	0.07	0.10	0.17
FE(BC-2)	-0.05	-0.05	0.01	0.07	0.08	0.05	0.11
FE(BC-3)	-0.07	-0.07	0.01	0.09	0.08	0.09	0.16

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009) with the number in the bracket indicating the bandwidth parameter.

Table 12: Correctly Specified Dynamic Probit; β

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
RE	0.01	-0.00	0.01	0.08	0.10	0.00	0.03
RE(R)	0.15	0.13	0.03	0.15	0.11	0.28	0.37
FE	0.57	0.56	0.38	0.57	0.22	0.74	0.92
FE(BC-0)	-0.25	-0.23	0.08	0.26	0.13	0.30	0.62
FE(BC-1)	-0.22	-0.21	0.07	0.23	0.15	0.24	0.49
FE(BC-2)	-0.27	-0.25	0.10	0.28	0.16	0.42	0.69
FE(BC-3)	-0.31	-0.29	0.12	0.32	0.16	0.60	0.86
T=8							
RE	0.01	0.00	0.00	0.05	0.06	0.02	0.05
RE(R)	0.02	0.01	0.00	0.05	0.06	0.04	0.08
FE	0.24	0.24	0.07	0.24	0.08	0.81	0.92
FE(BC-0)	0.00	0.00	0.00	0.05	0.06	0.01	0.04
FE(BC-1)	0.02	0.02	0.00	0.05	0.06	0.02	0.04
FE(BC-2)	0.02	0.02	0.00	0.05	0.06	0.02	0.04
FE(BC-3)	0.01	0.01	0.00	0.05	0.06	0.02	0.05
T=12							
RE	0.00	0.00	0.00	0.04	0.05	0.02	0.05
RE(R)	0.01	0.01	0.00	0.04	0.05	0.05	0.09
FE	0.15	0.15	0.03	0.15	0.06	0.75	0.86
FE(BC-0)	0.01	0.01	0.00	0.04	0.05	0.03	0.06
FE(BC-1)	0.02	0.02	0.00	0.04	0.05	0.03	0.06
FE(BC-2)	0.02	0.01	0.00	0.04	0.05	0.03	0.06
FE(BC-3)	0.01	0.01	0.00	0.04	0.05	0.02	0.06

500 replications. RE denotes the random effects estimator (Wooldridge, 2005). RE(R) denotes the random effects estimator performed on time-varying individuals. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009) with the number in the bracket indicating the bandwidth parameter.

Table 13: Correctly Specified Dynamic Probit; APE_y

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
FE	-0.88	-0.89	0.80	0.88	0.11	0.97	1.00
FE(BC-0)	-0.90	-0.89	0.82	0.90	0.11	0.99	1.00
FE(BC-1)	-0.35	-0.35	0.14	0.35	0.12	0.03	0.17
FE(BC-2)	-0.63	-0.62	0.41	0.63	0.13	0.52	0.89
FE(BC-3)	-1.09	-1.09	1.21	1.09	0.11	1.00	1.00
LPM	-1.01	-1.01	1.06	1.01	0.17	1.00	1.00
LPM(R)	-0.81	-0.81	0.70	0.81	0.22	0.94	0.97
T=8							
FE	-0.53	-0.54	0.29	0.53	0.08	1.00	1.00
FE(BC-0)	-0.52	-0.53	0.28	0.52	0.08	1.00	1.00
FE(BC-1)	-0.19	-0.19	0.04	0.19	0.09	0.21	0.42
FE(BC-2)	-0.18	-0.19	0.04	0.18	0.10	0.20	0.39
FE(BC-3)	-0.25	-0.25	0.07	0.25	0.10	0.46	0.64
LPM	-0.44	-0.44	0.21	0.44	0.11	0.99	1.00
LPM(R)	-0.36	-0.36	0.14	0.36	0.12	0.86	0.91
T=12							
FE	-0.38	-0.39	0.15	0.38	0.07	0.99	1.00
FE(BC-0)	-0.37	-0.38	0.14	0.37	0.07	0.99	1.00
FE(BC-1)	-0.12	-0.13	0.02	0.13	0.07	0.18	0.35
FE(BC-2)	-0.10	-0.10	0.02	0.11	0.08	0.11	0.23
FE(BC-3)	-0.12	-0.12	0.02	0.12	0.08	0.16	0.33
LPM	-0.25	-0.25	0.07	0.25	0.08	0.87	0.91
LPM(R)	-0.21	-0.21	0.05	0.21	0.09	0.70	0.79

500 replications. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009) with the number in the bracket indicating the bandwidth parameter. LPM denotes the linear fixed effect estimator. LPM(R) denotes the linear fixed effect estimator performed on time-varying individuals.

Table 14: Correctly Specified Dynamic Probit; APE_x

	Mean Bias	Median Bias	MSE	MAE	SD	p; .05	p; .10
T=4							
FE	-0.24	-0.24	0.06	0.24	0.06	0.02	0.22
FE(BC-0)	-0.31	-0.29	0.10	0.31	0.09	0.64	0.94
FE(BC-1)	-0.26	-0.25	0.08	0.26	0.09	0.26	0.62
FE(BC-2)	-0.34	-0.32	0.13	0.34	0.10	0.66	0.94
FE(BC-3)	-0.40	-0.39	0.17	0.40	0.09	0.92	0.99
LPM	-0.18	-0.18	0.04	0.18	0.07	0.77	0.85
LPM(R)	0.55	0.55	0.31	0.55	0.10	1.00	1.00
T=8							
FE	-0.15	-0.15	0.02	0.15	0.04	0.61	0.81
FE(BC-0)	-0.14	-0.14	0.02	0.14	0.04	0.72	0.88
FE(BC-1)	-0.12	-0.12	0.02	0.12	0.04	0.44	0.69
FE(BC-2)	-0.12	-0.12	0.02	0.12	0.04	0.46	0.71
FE(BC-3)	-0.13	-0.13	0.02	0.13	0.04	0.58	0.77
LPM	-0.13	-0.13	0.02	0.13	0.04	0.89	0.93
LPM(R)	0.19	0.19	0.04	0.19	0.05	0.97	0.99
T=12							
FE	-0.13	-0.13	0.02	0.13	0.03	0.87	0.94
FE(BC-0)	-0.13	-0.12	0.02	0.13	0.03	0.90	0.96
FE(BC-1)	-0.11	-0.11	0.01	0.11	0.03	0.79	0.92
FE(BC-2)	-0.11	-0.11	0.01	0.11	0.03	0.79	0.90
FE(BC-3)	-0.12	-0.12	0.01	0.12	0.03	0.82	0.93
LPM	-0.12	-0.12	0.02	0.12	0.03	0.95	0.97
LPM(R)	0.08	0.08	0.01	0.08	0.04	0.56	0.70

500 replications. FE denotes the fixed effects estimator. FE(BC) denotes the bias-corrected estimator proposed by Fernández-Val (2009) with the number in the bracket indicating the bandwidth parameter. LPM denotes the linear fixed effect estimator. LPM(R) denotes the linear fixed effect estimator performed on time-varying individuals.