Monte Carlo Evaluations of Common State Dependence Estimators

Eirik Eylands Brandsås

Master of Philosophy in Economics

Department of Economics

University of Oslo

May 2014

Monte Carlo Evaluations of Common State Dependence Estimators
Eirik Eylands Brandsås

http://www.duo.uio.no/
Print: Reprosentralen, Universitetet i Oslo

Preface

First I want to extend my gratitude to my main advisor Ph.D. Manudeep Bhuller at the Research Department of Statistics Norway. He has supported me from the first thesis proposal (which he quickly rejected) to the finished thesis. Drawing on his knowledge has been inspiring and extremely helpful. Furthermore the advice and feedback from Associate Professor Christian Brinch at the Norwegian Business School has been most valuable. Professor Wiji Arulampalam and Assistant Professor Alpaslan Akay provided me with well documented code for which I am very grateful. Secondly I would like to thank the ESOP research center at the University of Oslo for providing me with the ESOP scholarship and complimentary coffee.

Furthermore there are many individuals whom indirectly contributed to this thesis. They all deserve mentioning: Inga Hlíf Melvinsdóttir for her tremendous inspiration, motivation and for being the wonderful person she is. Kristrún Mjöll Frostadóttir for helping me survive Hagrannsóknir II. Otto S. Lillebø, Mathias Dahle Bryde-Erichsen and Sondre Seilen for excellent companionship at the University of Bergen. At the University of Oslo there are many who should be thanked, but first and foremost among them is Tone Hedvig Berg. Nicolai Ellingsen deserve thanks for his excellent feedback and corrections.

Abstract

This thesis represents an attempt to provide a deeper knowledge of the finite sample properties of some econometric methods used to estimate the magnitude of state dependence in binary choice dynamic panel models. These models are often applied in labor economics. The models I evaluate are the Heckman method, Wooldridge method and the linear probability model using Arellano-Bond instruments (Heckman, 1981a,b; Wooldridge, 2005; Arellano and Bond, 1991). By carefully designing appropriate Monte Carlo experiments I test the models' performance under different assumptions and different distributions of the error term, individual-specific fixed effects and explanatory variables.

The results indicate that the Heckman method is the most precise estimator in most cases, followed by the linear probability model. The Wooldridge method, while seldom the most accurate, is shown to be robust to violated assumptions. The linear probability model breaks down when the process includes an age-trended variable and the Heckman method breaks down when the explanatory variable is correlated with the individual-specific fixed effects. In most cases the three estimation methods display satisfactory performance. There are only modest performance gains from increasing the number of observed time periods.

Contents

Tables

1 Introduction

An established finding in the literature on labor market dynamics is that the rates of persistence in individuals' labor market state - for instance employment, poverty or welfare receipt - are very high. Heckman (1981a) distinguishes two sources of persistence in labor market histories. First, individuals differ in terms of observed and unobserved personal characteristics. Persistent individual characteristics such as low education or health problems may induce persistence in labor market outcomes, for instance recurring non-employment across periods. If left unaccounted for, observed or unobserved persistent individual characteristics induce *spurious state dependence* in labor market histories. Second, a past unemployment spell may itself have an effect on the probability of being unemployed today. For instance, past unemployment might lead to 'gaps' in a résumé, which potential employers might interpret as a negative signal of the applicants' unobserved productivity. This direct effect of a past state on the probability of being in the state in a later period is referred to as *true* or *structural state dependence*.

Much effort has been directed in the empirical literature to identify the two sources of persistence in labor market histories.[1] Common approaches for identifying state dependence involve estimation of dynamic binary choice panel data models with permanent unobserved heterogeneity. Unfortunately in maximum likelihood estimation we cannot consistently estimate such models with unrestricted individual-specific fixed effects due to the *incidental parameters problem*; for each individual we add to the sample the number of parameters to be estimated increases at a one-to-one rate (Neyman and Scott, 1948). The presence of unobserved heterogeneity across individuals is then typically accounted for by either conditioning on individual-specific fixed effects or integrating out the individual-specific fixed effects to get consistent estimates.

Unfortunately, these dynamic models still suffer from a range of identification problems. Binary choice panel data models suffer from the initial conditions problem (Heckman, 1981a). For instance, in random-effects probit models the unobserved individual-specific errors must be integrated out to construct a viable likelihood function. This requires one to specify the relationship between the individual-specific error and the outcome in the initial period, which enters the model as the lag of the outcome in the first observed period. Heckman (1981a) and Wooldridge (2005) propose solutions to the initial

[1] Examples from the existing literature are Chay et al. (1999) that study dynamics in welfare benefit receipt in the U.S., Stewart (2007) studies unemployment dynamics in the U.K., Biewen (2009) study state dependence in poverty in Germany and finally Bhuller and Brandsås (2013) study state dependence in poverty among immigrants in Norway.

conditions problem through alternative distributional assumptions on the relationship between the individual-specific error and the initial outcome. Both are commonly employed in the empirical literature. Meanwhile, the dynamic logit model has been proposed as an alternative that does not suffer from the initial condition problem (see e.g. Honoré and Kyriazidou (2000)), but comes with the cost of impeding calculation of marginal effects and strong restrictions on the exogenous covariates.

A simpler alternative to dynamic discrete-choice models is the dynamic linear probability model. Estimation of dynamic linear probability models with fixed effects using short panels can lead to substantial bias. Unobserved heterogeneity in such models is therefore typically accounted for through within-individual transformations. These transformation however induce a correlation between the lagged dependent variable and the error term. Anderson and Hsiao (1981, 1982), Arellano and Bond (1991) and Blundell and Bond (1998) propose solutions to this endogeneity problem that have been commonly employed in empirical work for continuous outcome variables. Other possible approaches for estimating linear or non-linear fixed-effects models include various bias-correction models, see for example Fernández-Val and Weidner (2013) for a bias corrected dynamic probit model.

Akay (2012) evaluates the finite sample properties of the coefficient estimates from the Heckman and Wooldridge methods by performing Monte Carlo experiments. He shows that the Wooldridge method performs satisfactory only for panels with more than five periods, while the Heckman method is suggested for panels of shorter durations. The Wooldridge method's weak performance is a result of a misspecification as shown in Rabe-Hesketh and Skrondal (2013). They find that the methods have virtually the same properties in all sample sizes when both are correctly specified. Arulampalam and Stewart (2009) also compare the coefficient estimates of the two methods, again finding that none of the methods dominate the other. Arellano and Bond (1991) provide Monte Carlo experiments showing that their proposed estimator for dynamic linear models performs satisfactory when the outcome variable is continuous.

Despite considerable evidence showing satisfactory performance of common dynamic panel data models in Monte Carlo experiments, there are certain issues that remain unexplored. First, results presented in any Monte Carlo study are subject to the choice of the benchmark model used to simulate the data. For instance, researchers typically use normally distributed simulated data for evaluating probit models and log-normally distributed data for logit models. Nonetheless, the true data-generating process remains unknown in empirical work and the distribution of the simulated error term will *a priori* favor one method over the others. Interest therefore lies in investigating the importance of benchmarking in Monte Carlo studies of dynamic panel data models. In particular, do any of the binary choice panel data models discussed above perform equally well under alternative choices of the benchmark model and are therefore robust to a misspecification of the error distribution?

Secondly, the existing literature has focused on the model's estimated coefficients. As is well known the coefficient estimates of binary choice models are difficult to interpret in a meaningful way. Effort is thus usually directed towards identifying the average marginal effect of the explanatory variables on the probability of observing a positive outcome in empirical work. When identifying state dependence we care about the effect of the lagged state on the probability of experiencing the state, that is the partial effect, and not the coefficient estimate by itself. Furthermore even if the coefficients are estimated precisely the models might give imprecise estimates of the average partial effects. I therefore focus primarily on the models' performance on estimating the parameters of interest in empirical research, the average partial effects.

The primary aim of this study is to contribute to the literature on dynamic panel data models by extending our knowledge of the finite sample properties of the linear probability model, the Heckman method and the Wooldridge method through Monte Carlo experiments. Data are simulated under alternative data-generating processes on which each model specification is estimated. The data-generating processes differ in the number of observed time periods, number of explanatory variables and distributions of both observed and unobserved variables. The analysis will shed light on largely neglected properties of the common estimators utilized to identify state dependence in labor market histories. Specifically the models are tested on their performance when their assumptions are violated in ways we can expect to occur in empirical settings.

The structure of the thesis is as follows: Section 2 derives the models mathematically and includes a treatment of both the incidental parameters and initial conditions problems, section 3 reviews previous simulation studies, section 4 details the simulation studies performed in the thesis, in section 5 I report summary statistics of the finite sample performance and finally in section 6 the results are discussed.

2 Econometric models

In the following sections I develop the three models for which I evaluate the finite sample performance. I start by presenting the mathematical notation and definitions that are used throughout the text. While precise notation is always important, I believe it to be key for the following models, as they can be very similar with important differences 'hidden' in the notation. This is especially true for the Wooldridge method, where published articles that have passed peer-review use misspecified and thus inconsistent models, likely due to unclear and non-standard notation (Rabe-Hesketh and Skrondal, 2013). I then develop the linear probability model, the Wooldridge method and the Heckman method before I provide a short overview of some alternative estimation methods.

2.1 Notation

Throughout the text vectors and matrices are denoted in a bold typeface. There are N observed individuals and the last observed period is T, so there are in total NT observations, $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$ and each individual is observed in each period without any missing variables. I use s to denote an unspecified initial time period where the process starts before the initial *observed* period $t = 1$. y_{it} is the state variable and is unity if an individual i is in the state in period t and equals zero else wise. $\boldsymbol{y}_i$ is a $T \times 1$ column vector, where T denotes the final observed time period. Thus $\boldsymbol{y}_i \equiv (y_{i1}, y_{i2}, \ldots, y_{iT})'$. $\boldsymbol{x}_i$ is a column vector containing all exogenous explanatory variables for an individual in all time periods, where the first element equals unity to accommodate for the intercept. With one exogenous explanatory variable $\boldsymbol{x}_i \equiv (1, x_{i1}, x_{i2}, \ldots, x_{iT})'$ and with κ covariates $\boldsymbol{x}_i = (1, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})' = (x_{1,i1}, x_{2,i1}, \ldots, x_{\kappa,i1}, \ldots, x_{1,iT}, x_{2,iT}, \ldots, x_{\kappa,iT})'$. The column vector $\boldsymbol{x}_i$ is of dimension $(T\kappa + 1) \times 1$. With κ explanatory variables the vector $\boldsymbol{x}_{it}$ refers to the value of all covariates in period t; $\boldsymbol{x}_{it} \equiv (1, x_{1,it}, x_{2,it}, \ldots, x_{\kappa,it})'$, with dimension $(\kappa+1) \times 1$. To refer to lagged variables we write $\boldsymbol{x}_{i,t-1}$, that is the values of $\boldsymbol{x}$ for individual i in the period immediately preceding period t.

The individual-specific fixed effect (often referred to as individual heterogeneity or just the fixed effect) is modeled through the variable c_i while the idiosyncratic error term is denoted u_{it}. $\boldsymbol{\beta}$ is a row vector consisting of the elements $\boldsymbol{\beta} \equiv (\beta_0, \beta_1, \ldots, \beta_\kappa)$ with dimension $1 \times (\kappa + 1)$. Typically unknown parameters are denoted with symbols from the Greek alphabet. I use $\mathrm{P}(\bullet)$, $\mathrm{E}(\bullet)$ and so on to denote probabilities, expectations or other statistical operators for the enclosed expression.

2.2 The identification problem

An observed empirical regularity is that the probability of unemployment in the next period is higher for those who are currently unemployed than for the employed. The central question is whether this persistence in unemployment is a result of personal characteristics, such as age, education or ability; or whether experiencing the state by itself increases the probability of future unemployment. Some factors are relatively easy to control for by adding variables that map the individuals' experience, age and other observable characteristics. But even after controlling for these variables (by including them in $\boldsymbol{x}_{it}$), individuals who have experienced poverty are more likely to experience it again, all else equal, as discussed in the introduction. There are two possible explanations for this empirical certainty (Heckman, 1981a).

The first explanation is that past experiences of a state alters behavior, preferences or constraints, which in part determine future outcomes of the state. Some intuitive, potential explanations for these alterations are reduced human capital due to unemployment, habit formation and reduced savings. Thus, experiencing the state in one period will affect the probability of experiencing the state in future periods. This is defined as true state dependence which is the parameter of interest in the thesis. If there is no true state dependence the state in the preceding period has no effect on the probability of experiencing the state, so that $y_{i,t-1}$ does not appear in (4). The effect of true state dependence will vary over time as individuals enter and leave the state.

Another potential cause for the observed persistence, after controlling for observable variables, is unobserved individual heterogeneity, where individuals differs in unobserved ways. We may expect that higher levels of education increase the probability of employment. Thus, if we do not control for education levels we may falsely claim that there exists true state dependence. However there exist other individual characteristics that we usually cannot observe, such as motivation, ability and social capital. Some of these unobserved factors are fixed over time, which we call unobserved permanent heterogeneity, captured by the individual-specific fixed effects c_i. The effect of the unobserved permanent heterogeneity varies between individuals as c_i varies between individuals.

The identification problem is for the remainder of the analysis defined as the problem of how to separate these two sources of persistence; the effects of true state dependence ($y_{i,t-1}$) from individual unobserved heterogeneity (c_i). I throughout assume that an appropriate dynamic model does allow us to distinguish true state dependence separately from spurious state dependence. As c_i is unobservable it is difficult to conceive methods that satisfactory control for its effects. We will see in the following sections that each estimation method proposes different solutions to this identification problem.

Hyslop (1999) considers a third potential driver of state dependence, that of transitory individual differences in the idiosyncratic error terms. These transitory differences between individuals lead to serial correlation in the error term, which will be another source of persistence. Models that account for serial correlation in dynamic binary choice models are uncommon in empirical work. This thesis attempts to evaluate the models that are commonly employed in the literature. Since these models do not account for serial correlation I design my experiments in a way that rules out serial correlation in the idiosyncratic error term.

2.2.1 Random or fixed effects

In modern panel data econometrics one of the crucial issues is whether one operates in the so called random or fixed effect framework. In the random effects framework c_i is uncorrelated with the structural variables. In the fixed effects framework c_i is allowed to be correlated with the structural variables. The random effects assumptions are:

$$\text{Cov}(y_{i,t-1}, c_i) = 0 \tag{1}$$

$$\text{Cov}(\boldsymbol{x}'_{it}, c_i) = 0 \tag{2}$$

These assumptions cannot be tested in empirical settings. But we know that the individual-specific fixed effect and the lagged dependent variable are correlated by construction. To see this note that the only way c_i and $y_{i,t-1}$ can be uncorrelated is if c_i has no effect on the individuals' outcomes in all periods. If they are uncorrelated there is no identification problem. As discussed this is unrealistic in most microeconomic applications as we cannot accurately measure ability or motivation, but we should acknowledge that it certainly affects the individual outcomes. A further complication arises as c_i is likely correlated with the strictly exogenous explanatory variables ($\boldsymbol{x}_{it}$). For example in labor market outcomes the unobservable individual-specific fixed effects such as ability are almost certain to affect the level of education that the individual possess. Such arguments are easily conceived in most microeconomic settings. As the assumption in (1) does not hold the models must devise a solution to the identification problem of separating the effect of c_i from $y_{i,t-1}$ on the response probability, as the random effects assumptions do not hold.

In empirical work these two assumptions cannot be tested as c_i is unobserved. In this thesis I simulate data and therefore specify the relationships between the variables. By carefully designing the processes I therefore ensure that there is no serial correlation in the error term and whether the assumptions in (1) and (2) do hold or not. A further advantage of simulation studies is that one can also control the presence of omitted time varying variables. In linear models we know that omitted variables lead to biased estimates if the omitted variable is correlated with the explanatory variables. In probit models we can get consistent estimates of the state dependence if the omitted variable is uncorrelated with the other explanatory variables and normally distributed. If the omitted variable is correlated with the explanatory variables we cannot consistently estimate the state dependence (Wooldridge, 2010, p 585). That we control the true underlying processes means that we *know* what the true distributions are and we therefore do not need to assume anything, we know whether the models' assumptions are true or not.

2.3 Response probability in binary choice models

In general, when we have binary models we want to find the probability of observing a given outcome instead of the actual outcome, realizing that the outcome, unity or zero, is a result of process that includes pure randomness. The randomness is modeled through the idiosyncratic error term u_{it}. Typically, we then formulate the response probability:

$$\mathrm{P}(y_{it} = 1|\boldsymbol{x}_i', y_{i,t-1}, \ldots, y_{i,0}, c_i) = F(\boldsymbol{x}_i', y_{i,t-1}, \ldots, y_{i,0}, c_i), \tag{3}$$

where F is an unspecified function, usually assumed to be the cumulative density function (CDF) of the error term u_{it}. Any outcome where $y_{it} = 1$ is called a success. Thus the probability of success is determined by the vector of explanatory variables, previous states and the individual-specific fixed effect. We will explore several different formulations of this general specification for F and its inputs. In the rest of section 2 we will assume that all variables in $\boldsymbol{x}_i$ are strictly exogenous conditional on c_i, and that there is only first order state dependence:

$$\mathrm{P}(y_{it} = 1|\boldsymbol{x}_i', y_{i,t-1}, \ldots, y_{i,0}, c_i) = \mathrm{P}(y_{it} = 1|\boldsymbol{x}_{it}', y_{i,t-1}, c_i) = F(\boldsymbol{x}_{it}', y_{i,t-1}, c_i) \tag{4}$$

Thus the response probability depends on the contemporaneous values of $\boldsymbol{x}_i$, the state in the preceding period and the individual-specific fixed effect. The structural variables are $y_{i,t-1}$ and $\boldsymbol{x}_{it}'$. That there is only first order state dependence means that there is no correlation between $y_{i,t-2}$ and y_{it} after conditioning on $y_{i,t-1}$.

2.4 Linear probability model

To develop the linear probability model (LPM) one can start by modeling the binary outcome as a linear function of the inputs:

$$y_{it} = \boldsymbol{\beta}\boldsymbol{x}_{it} + \rho y_{i,t-1} + c_i + u_{it} \tag{5}$$

Since y_{it} is a binary variable we know that the (conditional) expectation is the probability of success, which we use to to find the response probability:

$$\begin{aligned} \mathrm{E}(y_{it}|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) &= 1 \cdot \mathrm{P}(y_{it} = 1|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) + 0 \cdot \mathrm{P}(y_{it} = 0|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) \\ &= \mathrm{P}(y_{it} = 1|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) = F(\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) \end{aligned} \tag{6}$$

In the LPM the key assumption is that the response probability, determined by F, is a linear function of its inputs and furthermore that the idiosyncratic error term is uncorrelated with the other right hand side variables:

$$\mathrm{P}(y_{it} = 1|y_{i,t-1}, \boldsymbol{x}'_{it}, c_i) = \boldsymbol{\beta}\boldsymbol{x}_{it} + \rho y_{i,t-1} + c_i, \tag{7}$$

which can be straight forwardly estimated using standard panel data methods.

While the LPM results in very simple estimation procedures and simple inference it has some other issues. First, there are two minor drawbacks when using the LPM: (i) the error term is heteroskedastic and (ii) the error terms are not normally distributed. The variance of the error term can be expressed as $\mathrm{Var}(u_{it}|\boldsymbol{x}_{it}, y_{i,t-1}, c_i) = (1-[\boldsymbol{\beta}\boldsymbol{x}_{it}+\rho y_{i,t-1}+c_i])(\boldsymbol{\beta}\boldsymbol{x}_{it}+\rho y_{i,t-1} + c_i)$ which depends on the values of $\boldsymbol{x}_{it}$ and $y_{i,t-1}$, so it is heteroskedastic. From the same expression we see that the error term cannot be normally distributed. These two violations of the classical assumptions of OLS are minor as solutions exist using robust standard errors and/or feasible generalized least squares methods. Furthermore, even if one ignores the heteroskedasticity and non-normality of the error term, the problems do not affect consistency of the coefficient estimates, only the consistency of the estimated standard errors.

On the other hand there exists a more crucial problem with the LPM: it is almost always inconsistent, and usually biased, unless $\boldsymbol{\beta}\boldsymbol{x}_{it} + \rho y_{it} + c_i \in [0, 1]$ for all observations (Horrace and Oaxaca, 2006). Unfortunately there is little research done on the importance and size of the bias in the literature. Furthermore we know that the LPM can never be the true empirical model, unless further restrictions are placed on the idiosyncratic errors, as probabilities can exceed the possible range inside the unit interval: $\mathrm{P}(y_{it} = 1|y_{i,t-1}, \mathbf{x}_{it}, c_i)$ can be > 1 or < 0, a logical fallacy. This is seen by setting β_1 equal any positive non-zero value. Then, continuously increasing $x_{1,it}$ while holding the other variables constant will ensure that $\mathrm{P}(y_{it} = 1|y_{i,t-1}, \mathbf{x}_{it}, c_i) = 1$ at some value of $x_{1,it}$, and for even higher values of $x_{1,it}$ the probability exceeds 1.

There are some justifications for using the LPM; (1) there can be issues in binary choice models if you have endogenous variables that are easily handled in the LPM (2) as shown in section 2.4.2 it is easier to interpret the estimated coefficients as they give the marginal effects directly and (3) it is computationally easier. As we will see the first justification might be especially relevant in the current context where we want to estimate dynamic binary choice models.

2.4.1 Dynamics in the linear probability model

Estimating (7) by pooled OLS or the random effects estimator leads to biased estimates as the unobserved and omitted c_i is correlated with the other explanatory variables, as discussed in section 2.2. In panel data settings this problem is usually solved by transforming the variables by either first-differencing or within-transformation:

$$\Delta y_{it} = \boldsymbol{\beta}\Delta\mathbf{x}_{it} + \rho\Delta y_{i,t-1} + \Delta u_{it}, \tag{8}$$

where $\Delta u_{it} = u_{it} - u_{i,t-1}$. The usual within-transformation leads to the fixed effects estimation equation:

$$\bar{y}_{it} = \boldsymbol{\beta}\bar{\mathbf{x}}_{it} + \rho\bar{y}_{i,t-1} + \bar{u}_{it}, \tag{9}$$

where $\bar{u}_{it} = u_{it} - \frac{\sum_{t=1}^{T} u_{it}}{T}$, i.e. one subtracts the within-individual mean from each variable in each period. By transforming the data we have completely removed c_i, and any other time-constant variables, such as gender, from the equation of interest. In (8) and (9) the transformations have enabled us to get the estimation equations independent of c_i, and thus consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$ does not require the assumptions on zero correlation between the individual-specific fixed effect and the other explanatory variables, as in equations 1 and 2. A key concept to recognize is that while the estimation equation and variables are changed, the coefficients are the same. We can therefore estimate (8) or (9) to estimate the coefficients of interest from the linear response probability, (7).

At the same time as the transformations solves the identification problem it induces another problem: by construction $y_{i,t-1}$ and $u_{i,t-1}$ are correlated. Thus we have replaced the correlation problem between $y_{i,t-1}$ and c_i with another problem. In other words the usual solution to the correlation between the individual-specific fixed effect and other explanatory variables insert the lagged error term into the equation. The lagged error term is correlated with $y_{i,t-1}$. The solution to this problem was first proposed by Anderson and Hsiao (1981) for the first-differenced equation, where they proposed a pooled OLS estimation of (8) using $y_{i,t-2}$ or $\Delta y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$. As the estimation equation is first-differenced $u_{i,t-2}$ does not enter the equation and the instruments are uncorrelated with errors, assuming that the idiosyncratic error is serially uncorrelated. The method was later developed by Arellano and Bond (1991) in a generalized method of moments (GMM) where the set of instruments also include earlier lags of the dependent variables, based on the moment conditions:

$$E(y_{i,t-j}\Delta u_{it}) = 0, \quad \text{for } t = 3, \ldots, T \text{ and } j \geq 2, \tag{10}$$

which in total gives $(T-1)(T-2)/2$ orthogonality conditions that can be used as instruments. Several further developments have been proposed, such as the Blundell-Bond method (Blundell and Bond, 1998).[2] Note that the assumption in (10) does not hold if u_{it} is serially correlated. Then, if the error term is serially correlated one period back in time we must let $j \geq 3$, to avoid correlation between the instruments and the transformed idiosyncratic errors. Whether the errors are serially correlated or not can be tested with the Arellano and Bond (1991) test for serial correlation.

By first-differencing the data we eliminated the individual-specific fixed effect, and thus solved the identification problem. Secondly, eliminating c_i directly solves the correlation problem between c_i and the other explanatory variables. The crucial assumption for using the Arrelano-Bond method is that the idiosyncratic error terms are not serially correlated. This ensures that u_{it} is uncorrelated with the instruments and that the instruments have enough predictive power of the lagged dependent variable. For the remainder of the thesis LPM is the LPM with Arellano-Bond instruments.

2.4.2 Partial effects in the linear probability model

As mentioned above one of the advantages of the LPM is that it simplifies obtaining the partial effects greatly, compared to the other binary choice models. It is easily seen that the coefficient estimates are the partial effects of $x_{j,it}$ on the probability of success (assuming that there are no functional relationships between the covariates) by differentiating (8) with respect to $x_{j,it}$:

$$\frac{\delta P(y_{it} = 1 | \boldsymbol{x}'_{it}, y_{i,t-1}, c_i)}{\delta x_{j,it}} = \frac{\delta}{\delta x_{j,it}} (\boldsymbol{\beta x}_{it} + \rho y_{i,t-1} + c_i) = \beta_j \tag{11}$$

So the ceteris paribus effect of a one unit increase in $x_{j,it}$ leads to a β_j change in the probability of success. This partial effect is identical for all individuals, unless quadratics

[2] In the Blundell-Bond method one also use the first-differenced $\Delta y_{i,t-j}$'s as instruments, further increasing the set of valid instruments. In the thesis I will only employ the Arellano-Bond method. While the Blundell-Bond method is often used as an alternative to the Arellano-Bond method, it requires that the initial observed outcome is drawn from a steady state distribution for consistency. This, as I discuss in further detail in section 2.5.2, is unlikely to hold in the microeconomic applications where the dynamic binary outcome models are usually used. (Blundell and Bond, 1998). Furthermore the two methods in general give similar results, especially when ρ is not 'large' (Stewart, 2007).

and interactions are included. Thus the individuals' partial effect is also the average partial effect (APE).

2.5 General formulation of binary choice models

To develop the non-linear binary choice models we rely on an underlying latent variable model, where we let y^*_{it} be an latent continuous variable that depends on individual heterogeneity, the previous state, some strictly exogenous variables and a stochastic error term:

$$y^*_{it} = \boldsymbol{\beta x}_{it} + \rho y_{i,t-1} + c_i + u_{it} \tag{12}$$

The idiosyncratic error term follows an assumed known distribution with a known variance, σ^2_u. We postulate that if the latent variable is positive the individual experience the state:

$$y_{it} = 1\{y^*_{it} > 0\}, \tag{13}$$

where $1\{\bullet\}$ is an indicator function, taking unity if the enclosed statement is true and zero else wise. From (13) we find the conditional probability of success, i.e. the response probability:

$$\mathrm{P}(y_{it} = 1|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) = \mathrm{P}(y^*_{it} > 0|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) \tag{14}$$

This can be contrasted with the procedure in the LPM, where one begins by modeling the actual outcome, while one in the binary choice models specify a latent variable that determines the outcome. Importantly, both methods lead to a response probability.

Inserting for y^*_{it} in (14) gives:

$$\mathrm{P}(\boldsymbol{\beta x}_{it}+\rho y_{i,t-1}+c_i+u_{it} > 0|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) = \mathrm{P}(u_{it} > -\boldsymbol{\beta x}_{it}-\rho y_{i,t-1}-c_i|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) \tag{15}$$

We let $G(\bullet)$ denote the CDF of u_{it}. Then using the basic properties of CDFs and assuming that the probability density function (PDF) is symmetric about zero we get:

$$\mathrm{P}(y_{it} = 1|\boldsymbol{x}'_{it}, y_{i,t-1}, c_i) = 1 - G(-\boldsymbol{\beta x}_{it} - \rho y_{i,t-1} - c_i) = G(\boldsymbol{\beta x}_{it} + \rho y_{i,t-1} + c_i) \tag{16}$$

G is often referred to as the link function. If u_{it} is standard normally distributed we get $G = \Phi$, i.e. the CDF of standard normally distributed which leads to the probit model. The other common distributional assumption is the logistic distribution which leads to the logit estimator.

By letting the link function be a valid CDF we eliminate some of the problems connected to the LPM; chiefly that the probabilities cannot exceed the unit interval:

$$\lim_{\boldsymbol{\beta x}_{it}+\rho y_{i,t-1}+c_i \to -\infty} \mathrm{P}(y_{it}=1|y_{i,t-1},\boldsymbol{x}'_{it},c_i)=0 \tag{17}$$

$$\lim_{\boldsymbol{\beta x}_{it}+\rho y_{i,t-1}+c_i \to +\infty} \mathrm{P}(y_{it}=1|y_{i,t-1},\boldsymbol{x}'_{it},c_i)=1 \tag{18}$$

Unfortunately this comes at a cost; it is now harder to calculate the partial effects and misspecifying the link function generally leads to inconsistent coefficient estimates.

2.5.1 Partial effects in binary choice models

As is well known the estimated parameters in binary choice models can not generally be interpreted as the partial effect. Thus β_j does not measure the *ceteris paribus* effect of increasing $x_{j,it}$ by one unit on the probability of success. This is seen by differentiating (16) with regards to $x_{j,it}$:

$$\frac{\delta P(y_{it}=1|\boldsymbol{x}'_{it},y_{i,t-1},c_i)}{\delta x_{j,it}}=g(\boldsymbol{\beta x}_{it}+\rho y_{i,t-1}+c_i)\beta_j, \quad g(w)=\frac{\mathrm{d}G}{\mathrm{d}w}(w) \tag{19}$$

Unless G is linear the partial effects of a variable will depend on the other variables through $g(\bullet)$. Thus, there are several partial effects for each variable depending on which values of x_{it}, time periods and for which individuals one evaluates the partial effects at. The most common choice for dynamic models is the APEs. The APE of continuous variables is calculated by inserting for the observed values in (19) to get the partial effect for each individual in each time period and then averaging it across all individuals and time periods:

$$\mathrm{APE}(\beta_j)=\frac{\beta_j}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T} g(\boldsymbol{\beta x}_{it}+\rho y_{i,t-1}+c_i) \tag{20}$$

For binary and discrete variables the procedure is slightly different. For the APE of a binary variable one calculates the difference between (19) when the binary variable is equal to unity or zero, with the other variables held fixed at the observed values. To calculate the marginal effect of $y_{i,t-1}$ one thus evaluates:

$$\mathrm{APE}(\rho)=\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}[G(\boldsymbol{\beta x}_{it}+\rho+c_i)-G(\boldsymbol{\beta x}_{it}+c_i)] \tag{21}$$

Note that we only sum for periods $t=2,\ldots,T$ as y_{i0} is unobserved, and therefore we cannot estimate the partial effects for $t=1$. If $\boldsymbol{\beta},\rho$ and c_i are consistently estimated we get consistent estimates of the APEs by replacing the parameters with their estimated counterparts in (20) and (21).

A different method to evaluate the marginal effects is to calculate the partial effect at the average (PEA), where one insert for the average values of the observed variables. By inserting for different values of the covariates one can find other interesting measures, e.g. the average partial effect of $x_{j,it}$ at $y_{i,t-1} = 1$ and $y_{i,t-1} = 0$. In the literature estimating state dependence most researchers use APEs to evaluate the marginal effects.

Furthermore it is important to acknowledge the problem the unobserved c_i poses. Setting $c_i = C$ will only describe those individuals where this restriction holds. This is a significant problem with the binary choice models, which is completely avoided in the linear probability model. We will see that the Wooldridge and Heckman methods both solve this problem by placing a distributional assumption on c_i. In most studies with dynamic limited dependent variables state dependence, APE(ρ), is the main interest.

2.5.2 The incidental parameters and initial conditions problems

So far we have assumed strictly exogenous explanatory variables, first order state dependence and the distribution of u_{it}. The next step is to decide on how we treat the individual-specific fixed effects, c_i. One possibility is to treat the c_i's as parameters to be estimated, which leads to the so-called FE-probit estimator. The advantage of this method is that we avoid any assumptions on c_i and on the relationship between c_i and the other variables. Furthermore, we can then directly insert the estimated values of c_i into the formulas to calculate the partial effects. To estimate the binary-choice methods we use maximum likelihood and assume that $y_{i2}, \ldots, y_{iT}$ are independent conditional on $y_{i,t-1}, \boldsymbol{x}_{it}, c_i$. The conditional density which we base estimation on for individual i is:

$$\begin{aligned} f(y_{i2}, y_{i3}, \ldots, y_{iT}|y_{i1}, \boldsymbol{x}_i'; \boldsymbol{\beta}, \rho, c) &= \prod_{t=2}^{T} f_t(y_{it}|y_{i,t-1}, \boldsymbol{x}_{it}'; \boldsymbol{\beta}, \rho, c) \\ = \prod_{t=2}^{T} G(\rho y_{i,t-1} + \boldsymbol{\beta}\boldsymbol{x}_{it} + c)^{y_{it}} &[1 - G(\boldsymbol{\beta}\boldsymbol{x}_{it} + \rho y_{i,t-1} + c)]^{1-y_{it}}, \end{aligned} \tag{22}$$

where we treat c_i as a parameter to be estimated along with the structural parameters $\boldsymbol{\beta}$ and ρ. Note that the first observed state, y_{i1}, only appears as a conditioning variable, and that we do not evaluate the density in the first observed period, $t = 1$.

Unfortunately maximum likelihood estimation based on this conditional density leads to inconsistent estimates for all parameters (Neyman and Scott, 1948). Inconsistency arise because estimates of c_i are necessarily inconsistent when T is fixed as adding new individuals to the sample does not provide any additional information that allow us to determine c_i. Due to the non-linear nature of maximum likelihood the solution for the structural parameters involve the inconsistent estimates for the fixed effect which thus transmits the inconsistency. Simpler put, as we increase the number of individuals the number of parameters to be estimated increases at the same rate as we add individuals. This is the famous *incidental parameters* problem, first named in Neyman and Scott (1948). With fixed T asymptotics there is no log-likelihood that can be constructed that allow us to consistently estimate c_i, as we need $T \to \infty$.

The incidental parameters problem means that we cannot treat the individual-specific fixed effects as parameters to be estimated. This has an important implication when the ultimate goal of the analysis is the APEs and the degree of state dependence. As we have no estimates of c_i, we cannot consistently estimate the APEs without further assumptions.

To solve the problem we must in way specify the relationship between c_i and the other variables. In static models, where there are no lags of the dependent variable, the simplest solution is to assume that c_i is conditionally normally distributed:

$$c_i | \boldsymbol{x}_i' \sim \mathcal{N}(0, \sigma_c^2) \tag{23}$$

This assumption is unrealistic as it implies that c_i is independent of $\boldsymbol{x}_i'$. We can then integrate out c_i from the likelihood function, allowing ML-estimation of the other parameters. Unfortunately, with dynamics this raises the question on how to treat the initial observation y_{i1} and its relationship with c_i; the *initial conditions problem.* The simplest solution in dynamic models mimics the static random effects probit method. Keeping the assumption in (23) and assuming that y_{i1} is a non-stochastic starting position for the process we can integrate 22 against the density of c_i to obtain the density of $(y_2, y_3, \dots, y_T)$ which is not conditioned on c_i. The resulting density is then estimated by conditional maximum likelihood (CML) estimation.

Unfortunately this method an important drawback: That y_{i1} is non-stochastic implies that the individual fixed effect and the initial observed state are independent. Even in if we observe a process from the start, say employment history from graduation date for college graduates, the assumption is still unlikely to hold. The fixed effect almost certainly has an impact on the quality of college outcomes which again influences the initial state. The next two subsections develops two methods that give potentially consistent estimates of both $\boldsymbol{\beta}$ and ρ, the so-called Wooldridge and Heckman methods whom both propose different solutions to the initial conditions problem.

2.6 The Wooldridge method

The Wooldridge method was introduced by Wooldridge (2005), using a method similar to the one developed by Chamberlain (1980). Wooldridge's key insight is that by proposing certain densities for c_i it is possible to avoid conditioning on c_i, unlike in (22) where c_i was a conditioning variable. To do so, Wooldridge suggests to model the distribution of the unobserved effect conditional on the initial observed state and the exogenous variables in an auxiliary regression. The full Wooldridge method, as proposed and employed in the original paper is:

$$\mathbf{W}_1 : c_i = \gamma_1 y_{i1} + \boldsymbol{\gamma}_2 \boldsymbol{x}_i^{\dagger} + \zeta_i \tag{24}$$

Where $\boldsymbol{x}_i^{\dagger} = (1, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})'$ and $\zeta_i | y_{i1}, \mathbf{x}_i'^{\dagger} \sim \mathcal{N}(0, \sigma_\gamma^2)$ It is important to note that $\mathbf{x}_i^{\dagger}$ does *not* contain explanatory variables from the first observed period as Wooldridge (2005) shows that consistency requires that $c_i | y_{i1}, \boldsymbol{x}_i'^{\dagger}$ is correctly specified. But if the relationship between $\boldsymbol{x}_{i1}$ and c_i is strong we might get efficiency gains by including $\boldsymbol{x}_{i1}$ in the auxiliary regression. Later it has been shown that an estimator using $\boldsymbol{x}_i$ in the conditional density can be consistent (Rabe-Hesketh and Skrondal, 2013, p 347). We thus replace $\boldsymbol{x}_i^{\dagger}$ with $\boldsymbol{x}_i$ in $\mathbf{W}_1$:

$$\mathbf{W}_2 : c_i = \gamma_1 y_{i1} + \boldsymbol{\gamma}_2 \boldsymbol{x}_i + \zeta_i \tag{25}$$

As the number of variables in $\boldsymbol{x}_{it}$ and/or T grows both $\mathbf{W}_1$ and $\mathbf{W}_2$ will include a large number of variables which will reduce the degrees of freedom and complicates the integrals to be evaluated. This will increase estimation time significantly, even for relatively moderate panels. By constraining the effect of each element in $\boldsymbol{x}_{j,it}$ to be equal in all periods we can replace $\boldsymbol{x}_i^{\dagger}$ with $\overline{\boldsymbol{x}}_i^{\dagger}$ in $\boldsymbol{W}_1$: and $\boldsymbol{W}_2$, to get: $\overline{\boldsymbol{x}}_i = (\frac{\sum_{t=2}^{T} x_{1,it}}{T-1}, \frac{\sum_{t=2}^{T} x_{2,it}}{T-1}, \ldots, \frac{\sum_{t=2}^{T} x_{\kappa,it}}{T-1})$, and thus reduce the number of included variables. We thus get the following simplification for $\mathbf{W}_1$:

$$\mathbf{W}_3 : c_i = \gamma_1 y_{i1} + \boldsymbol{\gamma_2} \overline{\boldsymbol{x}}_i^{\dagger} + \zeta_i \tag{26}$$

And for $\mathbf{W}_2$:

$$\mathbf{W}_4 : c_i = \gamma_1 y_{i1} + \boldsymbol{\gamma_2} \overline{\boldsymbol{x}}_i^{\dagger} + \gamma_3 \boldsymbol{x}_{i1} + \zeta_i \tag{27}$$

Compared to $\mathbf{W}_1$ this approach constrains the effect of the exogenous variable, evaluated at $t = 2, \ldots, T$, to be identical for each period. Including $\overline{\boldsymbol{x}}_i$ instead of $\overline{\boldsymbol{x}}_i^{\dagger}$ leads to biased results as shown in Rabe-Hesketh and Skrondal (2013). In $\mathbf{W}_4$ we therefore include both $\boldsymbol{x}_i^{\dagger}$ and the initial observed outcomes $\boldsymbol{x}_{i1}$, without restrictions on γ_3.

As proposed in Wooldridge (2005) one can potentially include interactions between the initial observed state and the average of the covariates of $\boldsymbol{x}_i$:

$$\mathbf{W}_5 : c_i = \gamma_1 y_{i1} + \boldsymbol{\gamma_2} \overline{\boldsymbol{x}}_i^{\dagger} + \gamma_3 \boldsymbol{x}_{i1} + \gamma_4 y_{i1} \overline{\boldsymbol{x}}_i^{\dagger} + \zeta_i \tag{28}$$

This final specification is often used in empirical research citing possible efficiency gains, but its finite sample properties are unknown. We then have five potentially consistent estimators, given that the conditional distribution of c_i is correctly specified. A key assumption is that ζ_i is unrelated to the initial state, y_{i1}.

There are three potent research questions that can be answered by comparing $\mathbf{W}_1$ through $\mathbf{W}_5$: (1) Are there efficiency improvements by including the initial observations of the exogenous variables, (2) what, if any, are the losses we incur by using the simplified approaches and (3) the importance of including interactions between the initial state and time-varying covariates.

2.6.1 Likelihood function and partial effects

In this section I develop the Wooldridge method and construct the likelihood function estimation is based on, using specification $\mathbf{W}_1$ in (24). Estimation with the other variants of the conditional distribution of c_i is similarly developed. We formulate the following auxiliary regression:

$$c_i = \gamma_1 y_{i1} + \boldsymbol{\gamma}_2 \boldsymbol{x}_i^{\dagger} + \zeta_i, \tag{29}$$

where we assume that $\zeta_i | y_{i1}, \mathbf{x}_i^{\prime\dagger} \sim \mathcal{N}(0, \sigma_\gamma^2)$. Inserting the auxiliary regression for c_i in (12) give:

$$y_{it}^* = \boldsymbol{\beta} \boldsymbol{x}_{it} + \rho y_{i,t-1} + \gamma_1 y_{i1} + \boldsymbol{\gamma}_2 \boldsymbol{x}_i^{\dagger} + \zeta_i + u_{it} \tag{30}$$

Then assuming that $u_{it} | \boldsymbol{x}_i^{\prime\dagger}, y_{i,t-1}, \ldots, y_{i1}, \zeta_i \sim \mathcal{N}(0,1)$ immediately leads to the probit response probability, as shown in section 2.5:

$$\Phi(\boldsymbol{\beta} \boldsymbol{x}_{it} + \rho y_{i,t-1} + \gamma_1 y_{i1} + \boldsymbol{\gamma}_2 \boldsymbol{x}_i^{\dagger} + \zeta_i) \tag{31}$$

Leaving ζ_i as a conditioning variable instead of a parameter to be estimated leads the following conditional density for an individual (dropping the i subscript to conserve space):

$$\begin{aligned} f(y_2, \ldots, y_T | y_1, \boldsymbol{x}^{\prime\dagger}, y_{t-1}, \zeta; \boldsymbol{\beta}, \rho) = \prod_{t=2}^{T} \Big(& \big[\Phi\big(\boldsymbol{\beta} \boldsymbol{x}_t + \rho y_{t-1} + \gamma_1 y_1 + \boldsymbol{\gamma}_2 \boldsymbol{x}^{\dagger} + \zeta\big)\big]^{y_t} \\ & \times \big[1 - \Phi\big(\boldsymbol{\beta} \boldsymbol{x}_t + \rho y_{t-1} + \gamma_1 y_1 + \boldsymbol{\gamma}_2 \boldsymbol{x}^{\dagger} + \zeta\big)\big]^{1-y_t} \Big) \end{aligned} \tag{32}$$

By integrating (32) against the $\mathcal{N}(0, \sigma_\zeta^2)$ density we are able to remove ζ from the set of conditioning variables, allowing estimation by maximum likelihood while avoiding the initial conditions problem:

$$\begin{aligned} f(y_2, \dots, y_T | y_1, \boldsymbol{x}'^{\dagger}, y_{t-1}; \boldsymbol{\beta}, \rho) = \int_{\mathbb{R}} \Bigg(\prod_{t=2}^{T} \Big\{ \big[\Phi\big(\boldsymbol{\beta}\boldsymbol{x}_t + \rho y_{t-1} + \gamma_1 y_1 + \boldsymbol{\gamma}_2 \boldsymbol{x}^{\dagger} + \zeta\big)\big]^{y_t} \\ \times \big[1 - \Phi\big(\boldsymbol{\beta}\boldsymbol{x}_t + \rho y_{t-1} + \gamma_1 y_1 + \boldsymbol{\gamma}_2 \boldsymbol{x}^{\dagger} + \zeta\big)\big]^{1-y_t} \Big) (1/\sigma_\zeta) \phi(\zeta/\sigma_\zeta) \Bigg) \mathrm{d}\zeta \end{aligned} \tag{33}$$

If we define $\boldsymbol{w}_{it} \equiv (\boldsymbol{x}'_{it}, y_{i,t-1}, y_{i1}, \boldsymbol{x}'^{\dagger}_i)$ and a suitable coefficient vector $\boldsymbol{b}$ we can rewrite the conditional density:

$$\begin{aligned} f(y_2, \dots, y_T | y_1, \boldsymbol{x}'^{\dagger}, y_{t-1}; \boldsymbol{\beta}, \rho) = \int_{\mathbb{R}} \Bigg(\prod_{t=2}^{T} \Big\{ [\Phi(\boldsymbol{b}\boldsymbol{w}_{it} + \zeta)]^{y_t} \\ \times [1 - \Phi(\boldsymbol{b}\boldsymbol{w}_{it} + \zeta)]^{1-y_t}) (1/\sigma_\zeta) \phi(\zeta/\sigma_\zeta) \Bigg) \mathrm{d}\zeta, \end{aligned} \tag{34}$$

which has exactly the same structure as the standard random effects probit model. The integral can be evaluated using Gaussian-Hermite quadrature, a special method to approximate a definite integral which is often used for integrals over a normal distribution. This allows for easy estimation using standard software using $\boldsymbol{w}_{it}$ as the vector of explanatory variable in a panel probit model.

To find the estimated APE of a continuous variable evaluate:

$$\widehat{\mathrm{APE}}(\beta_j)^{\mathbf{W}_1} = \frac{\hat{\beta}_j^*}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \phi(\hat{\boldsymbol{\beta}}^* \boldsymbol{x}_{it} + \hat{\rho}^* y_{i,t-1} + \hat{\gamma}_1^* y_{i1} + \hat{\boldsymbol{\gamma}}_2^* \boldsymbol{x}_i^{\dagger}), \tag{35}$$

where the $*$ superscript indicate that the coefficients have been multiplied by $(1+\hat{\sigma}_\zeta^2)^{-1/2}$. Similarly we find the estimated APE of $y_{i,t-1}$ by:

$$\begin{aligned} \widehat{\mathrm{APE}}(\rho)^{\mathbf{W}_1} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \Bigg(\Phi(\hat{\boldsymbol{\beta}}^* \boldsymbol{x}_{it} + \hat{\rho}^* + \hat{\gamma}_1^* y_{i1} + \hat{\boldsymbol{\gamma}}_2^* \boldsymbol{x}_i^{\dagger}) \\ - \Phi(\hat{\boldsymbol{\beta}}^* \boldsymbol{x}_{it} + \hat{\gamma}_1^* y_{i1} + \hat{\boldsymbol{\gamma}}_2^* \boldsymbol{x}_i^{\dagger}) \Bigg) \end{aligned} \tag{36}$$

Thus, by specifying a reduced form for c_i (equations (24) to (28)) the Wooldridge method provides a simple to implement solution to both the incidental parameters problem and the initial conditions problem that is consistent.

2.7 The Heckman method

Heckman (1981b, p 188) presents an alternative solution. In contrast to the Wooldridge method Heckman's method approximates the reduced form equation for the initial observed state, y_{i1}, by a probit function. Here one should, if possible, include pre-sample information in the reduced form. In effect, Wooldridge's method incorporates the density $f(y_{i2}, \dots, y_{iT}|y_{i1}, \boldsymbol{x}_i^\dagger)$ while Heckman's method considers the density $f(y_{i1}, \dots, y_{iT}|c_i, \boldsymbol{x}_i)$ and then integrates out the fixed effect. The key difference lies in that the Heckman method consider the conditional density of the initial observed outcome, y_{i1} and conditions on c_i instead of conditioning on the initial observed outcome, as done in the Wooldridge method.

To develop the Heckman method we modify the latent variable for $t = 1$ and assume that $u_{it} \sim \mathcal{N}(0, 1)$ for $t = 1, 2, \dots, T$, as in the Wooldridge method. In the Wooldridge method the next step was to model c_i directly. The Heckman method instead models the initial state by a probit link function:

$$y_{i1}^* = \boldsymbol{\gamma} \boldsymbol{z}_{i1} + v_{i1} \tag{37}$$

$$y_{i1} = 1\{y_{i1}^* > 0\}, \tag{38}$$

where $\boldsymbol{z}_{i1}$ is a vector of exogenous covariates, including $\boldsymbol{x}_{i1}$ and if possible other pre-sample variables that do not belong in the latent variable for subsequent periods. The composed initial period error term is $v_{i1} = \pi c_i + u_{i1}$, where v_{i1} is allowed to be correlated with c_i but uncorrelated with u_{it} for $t \geq 2$. By not constricting π to equal zero the method allow for dependence between the initial period composite error term v_{i1} and the fixed effect c_i. Inserting for v_{i1} into (37) give

$$y_{i1}^* = \boldsymbol{\gamma} \boldsymbol{z}_{i1} + \pi c_i + u_{i1} \tag{39}$$

We treat $\boldsymbol{\gamma}, \pi$ as nuisance parameters to be estimated. The conditional distribution for the first period for individual i is thus:

$$f_1(y_{i1}|\boldsymbol{z}_{i1}', c_i; \boldsymbol{\gamma}, \pi) = \Phi\left(\boldsymbol{\gamma} \boldsymbol{z}_{i1} + \pi c_i\right)^{y_{i1}} \left[1 - \Phi\left(\boldsymbol{\gamma} \boldsymbol{z}_{i1} + \pi c_i\right)\right]^{1-y_{i1}} \tag{40}$$

For the remaining periods the conditional density is unchanged from the FE-probit conditional density, except for that c_i enters as a conditioning variable, not as a parameter to be estimated:

$$\begin{aligned} f(y_{i2}, \dots, y_{iT}|y_{i,t-1}, \boldsymbol{x}_i', c_i; \boldsymbol{\beta}, \rho) &= \prod_{t=2}^{T} f_t\left(y_{it}|y_{i,t-1}, \boldsymbol{x}_{it}', c_i; \boldsymbol{\beta}, \rho\right) \\ = \prod_{t=2}^{T} \Big(\Phi(\boldsymbol{\beta} \boldsymbol{x}_{it} + \rho y_{i,t-1} + c_i)^{y_{it}} & [1 - \Phi(\boldsymbol{\beta} \boldsymbol{x}_{it} + \rho y_{i,t-1} + c_i)]^{1-y_{it}} \Big) \end{aligned} \tag{41}$$

The next step is to combine the two conditional densities and to integrate out c_i. To integrate out c_i we must assume a distribution, which is usually assumed to be the standard normal distribution. Choosing the standard normal distribution allows us to evaluate the integral using Gaussian-Hermite quadrature, as in the Wooldridge specification. The resulting likelihood contribution is:

$$
\begin{aligned}
&f(y_{i1}, \ldots, y_{iT} | y_{i,t-1}, \boldsymbol{x}_i', \boldsymbol{z}_{i1}', c_i; \boldsymbol{\beta}, \rho) \\
&= \int_{\mathbb{R}} \left(f_1\left(y_{i1} | \boldsymbol{z}_{i1}', c_i; \boldsymbol{\gamma}, \pi\right) \prod_{t=2}^{T} f_t\left(y_{it} | y_{i,t-1}, \boldsymbol{x}_{it}', c_i; \boldsymbol{\beta}, \rho\right) \right) \phi(c_i) \mathrm{d}c
\end{aligned} \tag{42}
$$

Inserting for $f_1(\bullet)$ and $f_t(\bullet)$ in (42) the integral is:

$$
\begin{aligned}
&\int_{\mathbb{R}} \Bigg(\Phi\left(\boldsymbol{\gamma z}_{i1} + \pi\sigma_c c_1\right)^{y_1} \left[1 - \Phi\left(\boldsymbol{\gamma z}_{i1} + \pi\sigma_c c_1\right)\right]^{1-y_1} \times \\
&\prod_{t=2}^{T} \left\{ \Phi(\boldsymbol{\beta x}_{it} + \rho y_{i,t-1} + c_i)^{y_t} [1 - \Phi(\boldsymbol{\beta x}_{it} + \rho y_{i,t-1} + c_i)]^{1-y_t} \right\} \Bigg) \phi(c_i) \mathrm{d}c,
\end{aligned} \tag{43}
$$

which again results in a likelihood that can be evaluated using Gaussian-Hermite quadrature. As the likelihood is non-standard one cannot use standard probit estimation programs to estimate the likelihood, and the estimation procedure must usually be manually programmed. In Stata the user written program *redprob* (Stewart, 2006) maximizes the likelihood function without any need for programming. The APEs are calculated as in the Wooldridge case, replacing $\hat{\sigma}_\zeta^2$ with $\hat{\sigma}_c^2$ and using the correct reduced form. To estimate the APE of $y_{i,t-1}$ one evaluates:

$$
\widehat{\text{APE}}(\rho)^{\mathbf{H}} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Phi\big(\hat{\boldsymbol{\beta}}^* \boldsymbol{x}_{it} + \hat{\rho}^*\big) - \Phi\big(\hat{\boldsymbol{\beta}}^* \boldsymbol{x}_{it}\big) \right), \tag{44}
$$

where the coefficients are multiplied by $(1 + \hat{\sigma}_c^2)^{-1/2}$.

2.8 Other solutions

One possible solution is to simply ignore the incidental parameters problem and estimate the FE-probit model. The Wooldridge and Heckman methods both integrate out c_i to avoid having to estimate the fixed effect, while keeping either c_i or ζ_i as a conditioning variable. As we saw in section 2.5.2 the incidental parameters problem leads to inconsistent estimates of the structural parameters for fixed T and $N \to \infty$. Even with relatively high T the estimates appear to be heavily biased and performs significantly worse than the Wooldridge or Heckman methods for dynamic binary choice models (Fernández-Val and Weidner, 2013).

There also exists methods that attempt to remove and/or correct for the bias emerging from the incidental parameters problem. These models are called bias-correction models. Recently Fernández-Val and Weidner (2013) develop a bias-correction model for dynamic binary choice models that can be used to estimate state dependence with relatively good results. Their method work by directly manipulating the first order conditions that to be maximized in maximum likelihood methods. I do not include any of these models as no general estimation software has been developed and the dynamic bias-correction models have not yet been applied in empirical research.

There also exists a class of logit models that avoid the incidental parameters problem. Unfortunately these methods, such as the Honoré-Kyriazidou method place strict assumptions on the exogenous variables and prohibits calculation of marginal effects. These are therefore rarely employed in empirical work. But the Honoré-Kyriazidou method has one major advantage compared to the Heckman and Wooldridge method in that it does not place any assumptions on c_i.

3 Previous findings on finite-sample performance

Various Monte Carlo Experiments (MCEs) have been undertaken in earlier research. In this section I will give a brief overview of earlier research in thematic order. There is no standardized method to quantify the performance of estimators, and so the reported results and measurements vary. In the literature three measures of performance are common: (1) the root mean squared error (RMSE), (2) the mean absolute error (MAE) and (3) the average bias (AB). For a parameter η these are calculated as:

$$\text{RMSE}(\eta) = \sqrt{\text{MSE}(\eta)} = \sqrt{\frac{\sum_{g=1}^{G}(\hat{\eta}_g - \eta)^2}{G}} \tag{45}$$

$$\text{MAE}(\eta) = G^{-1}\sum_{g=1}^{G}|\hat{\eta}_g - \eta| \tag{46}$$

$$\text{AB}(\eta) = G^{-1}\sum_{g=1}^{G}\frac{\hat{\eta}_g - \eta}{\eta} \quad \text{measured in } \% \tag{47}$$

Each article that I review uses more than one data generating process (DGP), and so the number of reported results are quite large. To summarize the findings on estimated coefficients I have calculated the average of each reported measurement across specifications for each article. I use these averaged averages in the discussion. A description of all the previously employed DGPs in articles I review are included in table 1 below.

Table 1: Overview of earlier Monte Carlo experiments

Auth.[‡]	s	N	T	$x_{it} \sim$	$c_i \sim$	$u_{it} \sim$	ρ	β_1	β_0	y_{is} or y^*_{is}	Missing or note[obs]
HK	0	250–1000	4, 8	$\mathcal{N}(0, \pi^2/3)$	$\sum_{t=0}^{T} x_{it}$	log	$\frac{1}{4}, \frac{1}{2}, 1, 2$	1	0	$x_{i0}\beta + c_i + u_{i0}$	No
HK	0	250–4000	?	$\chi^2(1)$	$\sum_{t=0}^{T} x_{it}$	log	$\frac{1}{4}, \frac{1}{2}, 1, 2$	1	0	$x_{i0}\beta + c_i + u_{i0}$	No
HK	0	250–4000	?	$\mathcal{N}(0, \pi^2/3)$	$\sum_{t=0}^{T} x_{it}$	log	$\frac{1}{4}, \frac{1}{2}, 1, 2$	1	0	$x_{i0}\beta + c_i + u_{i0}$	Irrelevant x_{it}'s
HK	0	250–4000	4, 8	$\alpha_0(\xi_{it} + \alpha_1 + 0.1t)$	$\sum_{t=0}^{T} x_{it}$	log	$\frac{1}{4}, \frac{1}{2}, 1, 2$	1	0	$x_{i0}\beta + c_i + u_{i0}$	No
AS	-25	200–1000	3–8	$0.1t + 0.5x_{i,t-1} + {\varepsilon_{it}}^1$	$N(0, [\frac{1}{2}, 1, \frac{3}{2}])$	$\mathcal{N}(0,1)$	$\frac{1}{4}, \frac{1}{2}, \frac{3}{4}$	-1	4	$\beta_0 + \beta_1 x_{is} + c_i + u_{is}, x_{is} \sim U(-3,2)$	No[2]
AS	0	200	5	$0.1t + 0.5x_{i,t-1} + {\varepsilon_{it}}^1$	$N(0, ?(1))$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	-1	-1	$\sim N(-0.45, 1)^3$	Exogenous ${y_{is}}^2$
AS	0	200	5	$0.1t + 0.5x_{i,t-1} + {\varepsilon_{it}}^1$	$N(0, ?(1))$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	-1	-1	$-0.45 + rc_i + \sqrt{1-r^2}u_{i0};^3 r \in (\frac{1}{5}, \frac{4}{5})$	No[2]
RS	-25	200	3–20	$\mathcal{N}(0,1)$	$\mathcal{N}([0, \frac{1}{2}\overline{x}_i], 1)^4$	$\mathcal{N}(0,1)?$	$\frac{1}{2}$	1	0	Bern(0.5)	No
RS	-25	200	3–20	$(\chi^2(1) - 1/\sqrt{2})$	$\mathcal{N}([0, \frac{1}{2}\overline{x}_i], 1)^4$	$\mathcal{N}(0,1)?$	$\frac{1}{2}$	1	0	Bern(0.5)	No
RS	-25	200	3–20	$0.5x_{i,t-1} + \mathcal{N}(0,1)$	$\mathcal{N}([0, \frac{1}{2}\overline{x}_i], 1)^4$	$\mathcal{N}(0,1)?$	$\frac{1}{2}$	1	0	Bern(0.5), $x_{is} \sim \mathcal{N}(0,1)$	No
RS	-25	200	3–20	$0.1t + 0.5x_{i,t-1} + {\varepsilon_{it}}^1$	$\mathcal{N}([0, \frac{1}{2}\overline{x}_i], 1)^4$	$\mathcal{N}(0,1)?$	$\frac{1}{2}$	-1	4	Bern(0.5), $x_{is} \sim U(-3,2)$	No
Akay	0	200	3–20	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	1	0	$\mathcal{N}(0,1)$	Exogenous y_{i0}
Akay	-25	200	3–20	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	1	0	$\mathcal{N}(0,1)$	No
Akay	-25	200	3–20	$(\chi^2(1) - 1/\sqrt{2})$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	1	0	$\mathcal{N}(0,1)$	No
Akay	-25	200	3–20	$0.5x_{i,t-1} + \mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	1	0	$\mathcal{N}(0,1)$, $x_{is} \sim \mathcal{N}(0,1)$	No
Akay	-25	200	3–8	$0.1t + 0.5x_{i,t-1} + {\varepsilon_{it}}^1$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	-1	4	$\mathcal{N}(0,1)$, $x_{is} \sim \mathcal{N}(0,1)$	No
Akay	-25	200	3–8	$0.1t + 0.5x_{i,t-1} + \mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\frac{1}{2}$	-1	4	$\mathcal{N}(0,1)$, $x_{is} \sim \mathcal{N}(0,1)$	No
Heck	-25	100	3	$0.1t + 0.5x_{i,t-1} + {\varepsilon_{it}}^1$	$\mathcal{N}(0, [1,3])$	$\mathcal{N}(0,1)$	$\frac{1}{10}, \frac{1}{2}$	$\frac{-1}{10}, 0, 1$	0	$\beta_1 x_{i,t} + \rho + c_i + {u_{i,t}}^5$	No

All methods use one exogenous explanatory variable, $\kappa = 1$. All methods use either 500 or 1000 replications of each MCE, except Heckman who use 25 repetitions. A question mark indicates that the information is unclear or not provided.

[‡] HK, AS, RS, Akay and Heck refers to Honoré and Kyriazidou (2000), Arulampalam and Stewart (2009),Rabe-Hesketh and Skrondal (2013), Akay (2012) and Heckman (1981b), respectively.

[1] $\varepsilon_{it} \sim U(-0.5, 0.5)$

[2] They used the same x_{it} in all Monte Carlo repetitions.

[3] Values choses to give the same sample means of observed y's as in the benchmark experiment.

[4] Here $\overline{x}_i$ is the within-group mean, including all observed periods (from $t = 0$ to T. When $E[c_i] = \overline{x}_i$ x_{it} is endogenous.

[5] Unknown specification in the first period.

Heckman (1981b) compares the Heckman method with the FE-probit model using 11 different DGPs. He reports only the average estimated coefficient for each specification and lets $T = 3$ for the Heckman method and $T = 8$ for the FE-probit model which limits the comparability across models. The averaged ABs for the Heckman method are 6% and 30%, which is 1.6 and 0.15 times that of the FE-probit model, for β_1 and ρ respectively. The averaged RMSEs for the Heckman method are 0.4 and 0.3 times that of the FE-probit model, for β_1 and ρ respectively.[3]

Arulampalam and Stewart (2009) estimate the Heckman and Wooldridge methods, using 22 different DGPs. For the Wooldridge method they use specification $\mathbf{W}_1$. They find no significant differences between the two models. The averaged ABs for the Heckman method are 0.8% and -3%, which is 0.4 and 1.4 times that of the Wooldridge method, for β_1 and ρ respectively. The averaged RMSEs for the Heckman method are 0.99 and 0.97 times that of the Wooldridge method, for β and ρ respectively.

For panels of short duration, less than five periods, Akay (2012) find that the Heckman method performs better than the Wooldridge method. Rabe-Hesketh and Skrondal (2013) shows that this finding is a result of a misspecification of the Wooldridge method, as he attempt to use specification $\mathbf{W}_3$ but erroneously includes $\overline{\boldsymbol{x}}_i$ instead of $\overline{\boldsymbol{x}}_i^{\dagger}$. They evaluate the first four of Wooldridge's methods ($\mathbf{W}_1$ through $\mathbf{W}_4$) and the misspecified model in Akay (2012) using 40 different DGPs and find that the misspecified model overestimates ρ significantly compared to the correctly specified models. The averaged AB(ρ) of $\mathbf{W}_3$ is $-1,8\%$, which is 2 times that of the other correct specifications. The averaged RMSE are virtually identical across all models, including the misspecified. They conclude that $\mathbf{W}_4$ performs better than $\mathbf{W}_3$ and that there is no difference on performance between $\mathbf{W}_1$ and $\mathbf{W}_2$. For $T > 5$ the correct specifications has less than 1% AB. No studies have evaluated the performance of $\mathbf{W}_5$, even though it is often used in the literature.

Honoré and Kyriazidou (2000) compares the Honoré-Kyriazidou method to the FE-probit model, using 16 different DGPs. The averaged ABs for the Honoré-Kyriazidou method are 5% and -8%, which is 0.13 and 0.07 times that of the FE-probit model, for β_1 and ρ respectively. The averaged MAEs for the Honoré-Kyriazidou method is 0.25 and 0.2 times that of the FE-probit model, for β_1 and ρ respectively. Carro (2007) propose a new bias correction method; a modified maximum likelihood (MML) estimator that treat the fixed effect as a parameter to be estimated. He estimates the MML, FE-probit and the Honoré-Kyriazidou methods using 11 different DGPs. The FE-probit performs

[3] I used 'RMSE'(η) $= \sqrt{\hat{\bar{\eta}}_g - \eta}$ and 'AB'(η) $= \frac{\hat{\bar{\eta}}_g - \eta}{\eta}$, where $\hat{\bar{\eta}}$ is the average estimated coefficient as each $\hat{\eta}_g$ is not reported.

significantly worse than the other two methods. The averaged ABs of the MML method are 0.07 and 2.7 times that of the Honoré-Kyriazidou method for β_1 and ρ respectively. The averaged MAEs are 0.6 and 1.4 times that of the Honoré-Kyriazidou method for β_1 and ρ respectively. The MML method seems to estimate the β_1 more precisely and ρ less precisely than the Honoré-Kyriazidou method. Unfortunately no MCEs have been undertaken on the linear dynamic probability model to my knowledge. Several studies has evaluated the linear dynamic model, where the dependent variable is continuous, and it is shown that it performs satisfactorily, see for example Arellano and Bond (1991) and Behr (2003).

In total the earlier research have shown that the Wooldridge and Heckman methods perform well in finite samples when their necessary assumptions are satisfied. The preliminary research on bias-correction methods show that they perform better than the Honoré-Kyriazidou method, while allowing for calculations for APEs and placing less restrictive assumptions on the explanatory variables. There are no studies done on the dynamic LPM. To the authors knowledge no studies have evaluated the finite sample performance of the methods on estimated APEs.

4 Monte Carlo experiments

In this section I describe the various DGPs that I use to simulate data. Before discussing the individual experiments I will describe in detail how the simulations are carried out. As discussed in the introduction simulation studies are limited in their empirical relevance, as they will never reflect the real world and are useful only as far as their implications for real applications are valid. It is therefore important to generate DGPs with care so that they are relevant for empirical work, and preferably test the models' robustness to violated assumptions.

At the same time it is seldom clear how one should generate the DGPs to achieve these goals. For example the true DGPs for the different processes we care about, such as poverty and unemployment, are unknown. This difficulty must be tackled when one decides how to model the initial outcome, y_{is}: There cannot be any lags of the outcome, as $y_{i,s-1}$ does not exist yet since the process starts in period s so it is clear that the initial period outcome is generated differently than the subsequent periods. The initial outcome might be described by the individual characteristics without a lagged state variable, i.e. the first probability is determined by $\boldsymbol{\beta x}_{is} + c_i$ At the same time there might be presample variables that only affect the initial outcome, so that we should include a vector

$\boldsymbol{z}_{is}$ of these variables in the initial response probability. Alternatively the initial outcome might be determined more by chance than subsequent periods or even a pure stochastic realization. In short there is no general argument that can be made to conclude that one method is better or more 'true' than others. In the earlier research several methods have been employed to generate the initial period. I have chosen to model the initial outcome as the outcome of a Bernoulli trial with probability 0.5 in every DGP as it has been used in previous research and since a Bernoulli draw is a relatively simple mechanism.

The entire simulation procedure is performed in the statistical computer program Stata 13 in the following way: I first a random seed to ensure that the results are reproducible. I then start simulating the first DGP, $\mathbf{MCE}_{1.1}$, where I first specify the number of individuals ($N = 200$), the initial period ($s = -24$) and the number of periods for the current simulation, so that each individual is observed from s to T. I first generate the right hand side variables by drawing values for $\boldsymbol{x}_i$, the individual fixed effect and the idiosyncratic error. I then draw the initial outcome as a Bernoulli trial with probability 0.5. Finally $y_{i,s+1}$ is drawn as a Bernoulli trial with probability given the true response probability, $\mathrm{P}(y_{i,s+1} = 1) = G(\boldsymbol{\beta}\boldsymbol{x}_{i,s+1} + \rho y_{i,s} + c_i)$. I then repeat this step iteratively until period T. I then drop all observations preceding $t = 1$. After dropping observations I calculate the true average partial effects as defined in (20) and (21). I then drop the variables that are unobserved in empirical settings. Before the different models are estimated the dataset thus consists of NT observations, containing only y_{it} and $\boldsymbol{x}_{it}$. I then estimate: (i) the LPM with Arrelano-Bond instruments, (ii) the Heckman method and (iii) the five different Wooldridge specifications. I calculate the AB, MAE and RMSE of both the APEs and coefficient estimates, which I save to a separate dataset. This is done 500 times for each specification, where I draw new realizations for every repetition. The process is then repeated with a higher T, $T = 3, 4, 5, 8, 12, 20$ for each DGP, in line with the earlier literature. I then simulate the next DGP *without* resetting the seed. Thus the observed variables are new in every repetition for each DGP. In total there are 10 different DGPs with 6 different durations, where each duration is simulated 500 times, giving a total of 30,000 simulations and 210,000 regressions. In appendix D I have attached the code used for $\mathbf{MCE}_{1.1}$, and all the code is available on request.

4.1 Distributions of the time-varying variable x_{it}

I first specify three relatively simple DGPs. In these experiments I include one explanatory variable, so $\kappa = 1$. In each experiment I use different distributions for x_{it}. Three of the

DGPs are used in earlier studies cited above and are reproduced here to cross-validate the results. Furthermore evaluating the LPM on these processes is a novel contribution. In all four experiments both the individual-specific fixed effect and the idiosyncratic error are standard normally distributed.

For the first DGP I use the benchmark design from Rabe-Hesketh and Skrondal (2013). Here the model follows a dynamic probit model with a normally distributed explanatory variable. This DGP can be thought of as a benchmark model as it is simple and all estimation methods should have convincing performance.

$$\textbf{MCE}_{1.1} \quad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0,1) \\ y_{it}^* = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} \sim \mathcal{N}(0,1) \\ \beta_0 = 0 \quad \beta_1 = 1 \quad \rho = 0.5 & u_{it} \sim \mathcal{N}(0,1) \end{array}$$

The second DGP follows one of the methods proposed by Akay (2012). It is well known that experiments with normally distributed explanatory variables often find lower bias' than studies with non-normal distributions (Honoré and Kyriazidou, 2000). I let x_{it} be χ^2 distributed, with 1 degree of freedom, which I normalize to have mean zero and unity variance, to generate a non-normal explanatory variable.

$$\textbf{MCE}_{1.2} \quad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0,1) \\ y_{it}^* = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} = (\chi^2(1) - 1)/\sqrt{2} \\ \beta_0 = 0 \quad \beta_1 = 1 \quad \rho = 0.5 & u_{it} \sim \mathcal{N}(0,1) \end{array}$$

In the third MCE the strictly exogenous covariate follows a so-called Nerlove process, which includes a time trend, a lagged element and a stochastic error term drawn from the uniform distribution. Thus x_{it} approximate the age-trended variables that often arise in many microeconomic panels. This process is therefore often used in an attempt to increase the empirical relevance of the simulations. Note that the initial outcome of x_{it} is drawn from the uniform distribution and that the coefficients are changed from the other experiments. Furthermore, as T increases we see that fewer and fewer individuals will experience $y_{it} = 1$, and so there will be less variation in the outcome variable which might affect estimation performance.

$$\textbf{MCE}_{1.3} \quad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0,1) \\ y_{it}^* = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} = \frac{t+24}{10} + 0.5 x_{i,t-1} + U(\frac{-1}{2}, \frac{1}{2}) \\ \beta_0 = 4 \quad \beta_1 = -1 \quad \rho = 0.5 & u_{it} \sim \mathcal{N}(0,1) \quad x_{i,-24} \sim U(-3,2) \end{array}$$

The three specified DGPs will serve as benchmarks that I will use to compare the estimators' performance on relatively simple DGPs. By comparing estimation results from these DGPs with those from more complex DGPs I can test if the earlier findings on performance are a result of the simple processes used, and if the findings hold for more realistic scenarios. We know from section 3 that both the Heckman and Wooldridge methods perform well for $\mathbf{MCE}_{1.1}$ through $\mathbf{MCE}_{1.3}$.

4.2 Distributions of the idiosyncratic error u_{it}

The distribution of the idiosyncratic error term is important for consistency for all three methods. In particular both the Wooldridge and Heckman methods assume that the idiosyncratic error is standard normally distributed so that the response probability is determined by the CDF of the standard normal distribution, Φ. The LPM on the other hand does not assume a specific distribution on the idiosyncratic error, but the model require that it is serially uncorrelated and uncorrelated with $(\boldsymbol{x}'_{it}, y_{i,t-1})$. Thus, the consistency of the Wooldridge and Heckman models hinges on correctly specifying G, as $G(\boldsymbol{\beta x}_{it} + \rho y_{i,t-1} + c_i)$ is the probability of success. Secondly, the true distribution of the stochastic error term decides the form of G and g. Those are important in determining the true APEs, which is what we want to estimate. Notice that the three following DGPs are identical with $\mathbf{MCE}_{1.1}$ in all aspects except the distribution of the idiosyncratic error term.

In the literature on binary choice models it is often maintained that the theoretical consequences of misspecifying G are not very large (Cameron and Trivedi, 2005, p. 472). These arguments are based on the findings of Ruud (1983). He shows that if the mean of each structural variable, conditional on $\boldsymbol{\beta x}_{it} + \rho y_{i,t-1}$, is linear in $\boldsymbol{\beta x}_{it} + \rho y_{i,t-1}$ then misspecifying G leads to inconsistent estimates, where the estimates are scaled by a fixed factor. Therefore the ratio of the slope parameters is consistently estimated. This assumption holds, for example, if the variables are multivariate normal. There are four important reasons for why this finding is not of much comfort when we are most interested in state dependence: (i) The result only holds for slope parameters which do not include ρ, (ii) the APEs include the intercept and other dummy coefficients where Ruud's result does not apply, (iii) we are not interested in the ratio of the coefficients and (iv) as our model includes a binary variable $(y_{i,t-1})$ the variables cannot be multivariate normal. Furthermore when the $\boldsymbol{x}_{it}$'s are multivariate normal the LPM always consistently estimate the APEs, regardless of the form of g. It is thus interesting to see whether changing the distribution of u_{it} from the standard normal distribution to other distributions affects the method's performance.

The binary choice models cannot identify both σ_u^2 and the coefficients of the latent variable, see appendix A. I therefore use two distributions for u_{it} with variation equal to unity to ensure comparability with the first four MCEs. In the third distribution the idiosyncratic error follows a logistic distribution. While the logistic distribution has higher variation estimating a DGP based on the logistic distribution provides relevant information: If the estimates still have decent properties when the error term is logistically distributed this means that the Wooldridge and Heckman methods may be suitable even when the probit normalization is false. Secondly as discussed the Honoré-Kyriazidou method assumes that the stochastic error term is logistically distributed. If the Heckman and Wooldridge methods have good performance even when the error term is logistically distributed this implies that we can use these methods even when we know that the true distribution is logistic. We can then avoid the restrictions that the Honoré-Kyriazidou method place on the explanatory variables. In all the following processes I only vary the distribution of the error term, both the individual fixed effect and the explanatory variable is standard normally distributed in all DGPs and the coefficients are set to $\beta_0 = 0, \beta_1 = 1, \rho = 0.5$.

In the first DGP the idiosyncratic error is a standardized Laplace-distribution, where I standardize the distribution with scale parameters $\sqrt{0.5}$ so that the variance equals unity and location 0 so that the mean equals zero. The Laplace distribution differs from the standard normal distribution by having fatter tails and a more distinct peak at the mean, while still being a symmetric and S-shaped CDF.

$$\textbf{MCE}_{2.1} \qquad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0,1) \\ y_{it}^* = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} \sim \mathcal{N}(0,1) \\ \beta_0 = 0 \quad \beta_1 = 1 \quad \rho = 0.5 & u_{it} \sim Laplace(0, \sqrt{0.5}) \end{array}$$

In the second DGP I use a uniform distribution with the minimum parameter set to $\frac{-\sqrt{12}}{2}$ and the maximum parameter set to $\frac{\sqrt{12}}{2}$. These parameter values ensure that the mean and variation equals zero and unity, respectively. The uniform distribution is an illogical choice for an actual empirical DGP for binary outcomes. It can shed light on the performance of the binary choice models when the assumed distribution is significantly different from the correct one.

The last distribution I employ is the exponential distribution. This distribution leads to the logit model. It is well known that the estimated coefficients of probit and logit

$$\textbf{MCE}_{2.2} \quad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0,1) \\ y_{it}^* = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} \sim \mathcal{N}(0,1) \\ \beta_0 = 0 \quad \beta_1 = 1 \quad \rho = 0.5 & u_{it} \sim \text{U}(\frac{-\sqrt{12}}{2}, \frac{\sqrt{12}}{2}) \end{array}$$

models in cross sections are comparable, after correcting for the different normalization. Even without correcting for the different scalings the average partial effects are usually very similar (Wooldridge, 2010, pp. 581,600-601). To test whether this result hold in dynamic panels I therefore generate logistically distributed error terms. Furthermore this also checks the effect of heavier tails and the effect of using the wrong scaling on the coefficient estimates on the estimated APEs.

$$\textbf{MCE}_{2.3} \quad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0,1) \\ y_{it}^* = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} \sim \mathcal{N}(0,1) \\ \beta_0 = 0 \quad \beta_1 = 1 \quad \rho = 0.5 & u_{it} \sim log(0,1) \end{array}$$

In figure 1 I have plotted the CDF and PDFs of the standard normal, Laplace, standard logistic and uniform distributions with parameters as defined. Notice that the differences between the distributions are larger in the PDFs that determine the APE of continuous variables than in the CDFs that determine the APE of the lagged dependent variable.

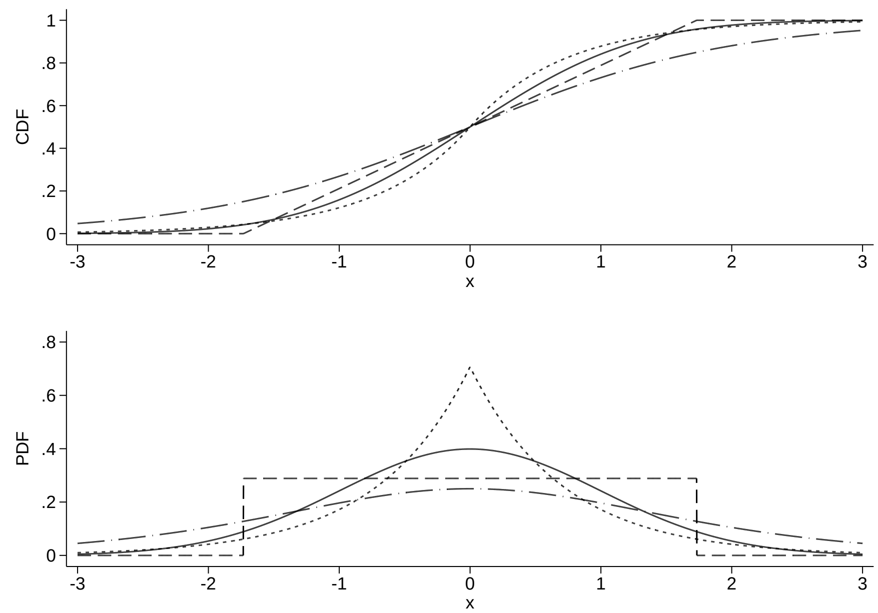

Figure 1: The CDF and PDF of the employed distributions. They are the standard normal (solid line), uniform (dashes), logistic (long dashes with dots) and Laplace (dotted).

4.3 Distributions of c_i, endogeneity and omitted variables

The previous three DGPs all use different distributions for the idiosyncratic error, while the first three all use different distributions on the single explanatory variable. I design the next three DGPs to allow me to do some preliminary research into the robustness of the models to: (i) different distributions for the individual-specific fixed effect, (ii) endogeneity by letting $\text{Cov}(x_{it}, c_i) \neq 0$ and (iii) correlated omitted variables.

As discussed in section 2 the LPM model does not place any assumptions on the individual-specific fixed effect, while the Heckman and Wooldridge methods both assume that it is standard normally distributed. No studies have evaluated their performance for other distributions of the fixed effect. Another potential cause of bias is the usual omitted variable bias. In general omitted variables that are correlated with the included explanatory variables will affect the coefficient estimates of the model and their estimated partial effects. In all of the following processes all DGPs run for 25 periods before they are observed, the initial outcome is generated as a binomial draw with probability 0.5 and stochastic error term is normally distributed. I only vary either the distribution of the individual-specific fixed effect or the strictly exogenous explanatory variable.

In the first DGP I let c_i equal a gamma distributed variable with shape 4 and scale 0.5, where I subtract 2 to get an individual-specific fixed effect with mean zero and unity variance.[4] Note that the variance and mean is unchanged from all earlier DGPs where c_i was generated as a standard normally distributed variable. The Wooldridge method assumes that ζ_i is standard normally distributed while the Heckman method assumes that c_i is standard normally distributed. Both assumptions are clearly violated with the gamma distributed c_i. The change in distribution should not affect the LPM as the first-differencing removes c_i from the estimation equation. Except for c_i this DGP is identical to $\mathbf{MCE}_{1.1}$.

$$
\begin{array}{lll}
 & y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \Gamma(4, 0.3) - 2 \\
\mathbf{MCE}_{3.1} & y_{it}^{*} = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} \sim \mathcal{N}(0,1) \\
 & \beta_0 = 0 \quad \beta_1 = 1 \quad \rho = 0.5 & u_{it} \sim \mathcal{N}(0,1)
\end{array}
$$

In the second trial I let x_{it} be normally distributed with mean $c_i/3$. Then c_i and x_{it} has correlation coefficient of 0.315 irrespective of T.[5] Again this should not affect the

[4]Letting X denote the gamma distributed variable and using the standard formulas for gamma distributed variables we get: $\text{E}(c_i) = E(X-2) = \alpha\lambda - 2 = 0$ and $\text{Var}(c_i) = Var(X-2) = \alpha\lambda^2 = 1$

[5]Using the method that Rabe-Hesketh and Skrondal (2013) employ leads to an endogeneity that decreases as T increases. For $T = 20$ the correlation coefficient is 0.01-0.05 in their specification.

LPM as c_i is eliminated in the first-differencing of the estimation equation. On the other hand the two other methods operate in the random effects framework where correlation between the covariates and the individual specific fixed effect can lead to inconsistency. Endogenous covariates are certainly relevant: as discussed in section 2.2 we expect that c_i is correlated with $\boldsymbol{x}_{it}$ in empirical microeconomic applications. Except for x_{it} and its correlation with c_i this DGP is identical to $\mathbf{MCE}_{1.1}$:

$$\mathbf{MCE}_{3.2} \qquad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0,1) \\ y_{it}^* = \beta_0 + \beta_1 x_{it} + \rho y_{i,t-1} + c_i + u_{it} & x_{it} \sim \mathcal{N}(\frac{c_i}{3}, 1) \\ \beta_0 = 0 \quad \beta_1 = 1 \quad \rho = 0.5 & u_{it} \sim \mathcal{N}(0,1) \end{array}$$

In the third specification I expand the true DGP to include two strictly exogenous covariates, $\mathbf{x}_{it} = (x_{1,it}, x_{2,it})$. They are bivariate normally distributed, with mean zero, unity variance and a correlation coefficient of 0.3. By omitting the second explanatory variable in the regression I can shed some preliminary light on the behavior of the estimators with omitted variables, a setting that is likely to be encountered in empirical research. Again $\mathbf{MCE}_{1.1}$ serves as a benchmark model as the only difference between it and $\mathbf{MCE}_{3.3}$ is the additional variable that is omitted in the regressions.

$$\mathbf{MCE}_{3.3} \qquad \begin{array}{ll} y_{i,-24} \sim \text{Bern}(0.5) & c_i \sim \mathcal{N}(0.1) \\ y_{it}^* = \boldsymbol{\beta} \boldsymbol{x}_{it} + \rho y_{i,t-1} + c_i + u_{it} & \boldsymbol{x}_{it} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \frac{3}{10} \\ \frac{3}{10} & 1 \end{bmatrix}\right) \\ \boldsymbol{\beta} = (0,1,1) \quad \rho = 0.5 & u_{it} \sim \mathcal{N}(0,1) \end{array}$$

5 Results

The results of $\mathbf{MCE}_1$, $\mathbf{MCE}_2$ and $\mathbf{MCE}_3$ are summarized in table 2, 3 and 4 respectively. The results are summarized for the LPM, the Heckman method and the fourth Woolridge specification, $\mathbf{W}_4$. The results for the four remaining Wooldridge methods are reported in appendix B. I include only one Wooldridge specification in the main text since the finite-sample results are similar across the different specifications of the Wooldridge method and to conserve space. For a further discussion of the differences between the Wooldridge specifications see appendices B and C. In all tables I report the AB measured in percentage points, the MAE and the RMSE for the estimated APEs. There are several reasons for the focus on APEs instead of the coefficient estimates, a substantial departure from the earlier literature: (i) In empirical work the focus lies primarily in the estimated APEs and seldom in the coefficient estimates themselves, (ii) it is a novel contribution to the field and (iii) the LPM model cannot identify the coefficients of the latent variable. In appendix C I report the average bias, RMSE and MAE of the coefficient estimates. I do not focus on the estimates of σ_c^2 as it is irrelevant for the estimated APEs of the LPM. Furthermore, even while σ_c^2 affects the estimated APEs (see equations 35 and 36) of the Heckman and Wooldridge method the measure of interest is generally the APEs of these methods, and not the estimated variation of the individual-specific fixed effect. For each table of results I first describe the results of each specific MCE (e.g. $\mathbf{MCE}_{1.1}$) before summarizing the implications of the finite sample results for each class of MCEs (e.g. $\mathbf{MCE}_1$). In section 5.4 I provide a short discussion of the coefficient estimates.

5.1 Finite sample results based on $\mathbf{MCE}_1$

In the first class of MCEs the processes are relatively simple and we expect good results. In the earlier literature these MCEs have been used to evaluate the finite sample performance of the Wooldridge and Heckman methods when estimating the coefficients of the latent variable, where they both have displayed good performance. Thus we expect the estimated APEs to also be precisely estimated for these two models. The LPM model has not yet been evaluated using these DGPs.

When evaluating the APE estimates based on the first DGP, $\mathbf{MCE}_{1.1}$, there are several striking patterns. When $T = 3$ the LPM overestimates the effect of state dependence, APE(ρ), with 23.01%, the Heckman method underestimates with -2.21% and the Wooldridge method overestimates with 8.44%. As T increase we see a decline in the average bias of the LPM, now ranging from 1.59% to 5.36%, while the Heckman method remains unbiased with ABs ranging from -0.83% to 3.53% and the Wooldridge method has ABs ranging from 5.26% to 9.50%. None of the methods display any significant reduction in the AB as the number of observed time periods increases. The ABs of APE(β_1) range from -0.35% to 1.24% for the LPM, -0,52% to 5.57% for the Heckman method and from 7.95% to 9.25% for the Wooldridge method. Overall the Heckman method displays convincing performance with ABs ranging from -2.21% to 5.57%. For $T > 3$ the LPM also performs well. The Wooldridge method overestimates the ABs with approximately 8%. An interesting result is that all three methods have more precise estimates for APE(β_1) compared to the APE(ρ) as measured by MAE and RMSE. All methods have decreasing MAEs and RMSEs as T increases, even when the ABs do not decrease. For every T

the Heckman method performs better than the other two methods. The LPM displays promising performance for $T > 3$.

Based on $\mathbf{MCE}_{1.2}$ a similar pattern is observed. For $T = 3$ the LPM overestimates APE(ρ) with 23.01% and for $T > 3$ the ABs range from 2.46% to 8.59%. The Heckman method again precisely estimate the APE with ABs ranging from 0.22% to 4.54% while the ABs of the Wooldridge method range from 8.50% to 12.40%. Thus we see that a non-normal distribution for x_{it} increase the ABs of the estimated APE(ρ) for all three models. The non-normality of x_{it} has even larger effects on the ABs of the estimated APE(β_1)s with ABs ranging from -31.28% to -29.5% for the LPM, 0.85% to 9.19% for the Heckman method and 8.37% to 12.35% for the Wooldridge method. The ABs of $\widehat{\text{APE}}(\beta_1)$ actually increase with T for both the Heckman and Wooldridge methods. The MAEs and RMSEs tend to decrease with T, with values only slightly higher than in $\mathbf{MCE}_{1.1}$, except for the LPM's estimates of APE(β_1). We can thus conclude that changing x_{it} to be a standardized χ^2 variable has only marginal impacts on $\widehat{\text{APE}}(\rho)$, but larger impacts on $\widehat{\text{APE}}(\beta_1)$, and especially so for the LPM.

The estimated APE(ρ)s from $\mathbf{MCE}_{1.3}$ deviate from the two earlier methods. The ABs of the LPM now range from 81.19% to 516.48%, and they are rapidly increasing in T. The Heckman method also has higher ABs, now ranging from 5.67% to 12.66% and the Wooldridge has ABs ranging from -3.26% to 16.83%. The LPM is heavily biased when estimating state dependence in a Nerlove process. As discussed in section 2.4 the LPM is inconsistent if $\boldsymbol{\beta x}_{it} + \rho y_{it} + c_i \notin [0, 1]$, a condition that is irrelevant for the Wooldridge and Heckman methods. For the Nerlove process 81.5% and 95.1% of observations, for $T = 3$ and $T = 20$ respectively, are outside the unit interval. For APE(β_1) the LPM performs better with ABs ranging from -2.35% to 33.53% compared to estimates ranging from -3.78% to 4.63% for the Heckman method and -3.34% to 4.77% for the Wooldridge method. We see the same pattern in the MAEs and RMSEs, for example the MAEs of $\widehat{\text{APE}}(\rho)$ range from 0.10 to 0.22 in the LPM estimates compared to 0.01 to 0.09 for the Heckman method and 0.01 to 0.10 in the Wooldridge method. The Wooldridge and Heckman method both have decreasing MAEs and RMSEs in T for both APEs, in similar intervals as for the two preceding MCEs.

Table 2: Finite sample APE results for $\mathbf{MCE}_1$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Linear probability model						Heckman						Wooldridge 4 ($\mathbf{W}_4$)					
	ρ			β_1			ρ			β_1			ρ			β_1		
T	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE
	$\mathbf{MCE}_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$																
3	23.01	0.15	0.19	1.24	0.02	0.03	-2.21	0.06	0.07	-0.52	0.03	0.03	8.44	0.07	0.09	8.23	0.03	0.03
4	1.59	0.08	0.10	0.07	0.02	0.02	1.76	0.04	0.05	2.50	0.02	0.02	9.50	0.05	0.06	7.95	0.02	0.03
5	5.36	0.06	0.07	-0.35	0.01	0.02	0.84	0.04	0.05	3.62	0.02	0.02	7.28	0.04	0.05	8.28	0.02	0.02
8	4.41	0.03	0.04	0.14	0.01	0.01	0.22	0.02	0.03	4.16	0.01	0.02	7.37	0.02	0.03	9.14	0.02	0.02
12	1.65	0.02	0.03	-0.05	0.01	0.01	-0.83	0.02	0.02	4.54	0.01	0.01	5.26	0.02	0.02	9.25	0.02	0.02
20	4.57	0.02	0.02	-0.05	0.01	0.01	3.53	0.01	0.02	5.57	0.01	0.01	7.69	0.01	0.02	9.20	0.02	0.02
	$\mathbf{MCE}_{1.2}$	$x_{it} \sim \frac{\chi_1^2 - 1}{\sqrt{2}}$																
3	23.01	0.15	0.20	-29.50	0.07	0.08	2.52	0.07	0.09	0.85	0.04	0.05	12.40	0.09	0.11	8.37	0.04	0.04
4	8.59	0.09	0.11	-30.00	0.07	0.07	2.09	0.05	0.06	4.88	0.03	0.04	12.13	0.05	0.07	10.82	0.03	0.04
5	2.46	0.06	0.08	-31.28	0.07	0.07	1.13	0.04	0.05	4.93	0.03	0.03	9.38	0.04	0.05	10.07	0.03	0.03
8	4.89	0.04	0.04	-31.23	0.07	0.07	0.22	0.03	0.03	6.09	0.02	0.02	8.50	0.03	0.03	11.10	0.03	0.03
12	6.68	0.03	0.03	-31.04	0.07	0.07	1.30	0.02	0.02	6.75	0.02	0.02	8.78	0.02	0.03	11.34	0.03	0.03
20	6.94	0.02	0.02	-30.41	0.07	0.07	4.54	0.01	0.02	9.19	0.02	0.02	8.94	0.02	0.02	12.35	0.03	0.03
	$\mathbf{MCE}_{1.3}$	x_{it} ~Nerlove-process																
3	81.19	0.15	0.19	-2.35	0.07	0.09	8.00	0.09	0.11	-3.78	0.05	0.07	15.77	0.10	0.12	1.53	0.07	0.09
4	83.26	0.11	0.14	-0.40	0.05	0.06	17.75	0.07	0.08	1.07	0.03	0.04	16.83	0.06	0.08	2.76	0.04	0.05
5	89.71	0.10	0.12	3.26	0.04	0.05	11.43	0.05	0.06	4.63	0.03	0.04	9.94	0.04	0.06	4.77	0.03	0.04
8	123.87	0.11	0.12	19.97	0.03	0.04	5.67	0.03	0.03	4.41	0.02	0.02	0.86	0.02	0.03	2.48	0.01	0.02
12	205.19	0.14	0.15	33.53	0.04	0.04	10.40	0.02	0.02	2.78	0.01	0.01	-0.58	0.02	0.02	-1.28	0.01	0.01
20	516.48	0.22	0.22	21.57	0.02	0.02	12.66	0.01	0.01	3.37	0.01	0.01	-3.26	0.01	0.01	-3.34	0.00	0.01

For detailed description of each MCE see section 4.1. $\mathbf{W}_4$ converged in each instance. Heckmans method failed to converge in 15 instances for $\mathbf{MCE}_{1.1}$, 23 instances for $\mathbf{MCE}_{1.2}$ and 274 instances for $\mathbf{MCE}_{1.3}$. The response probability outside the unit interval in 74.7% of repetitions in $\mathbf{MCE}_{1.1}$, in 73.3% of repetitions in $\mathbf{MCE}_{1.2}$ and in 87.5% of repetitions in $\mathbf{MCE}_{1.3}$. The averaged empirical APE(ρ) is 0.115 in $\mathbf{MCE}_{1.1}$, 0.121 in $\mathbf{MCE}_{1.2}$ and 0.087 in $\mathbf{MCE}_{1.3}$. The averaged empirical APE(β_1) is 0.218 in $\mathbf{MCE}_{1.1}$, 0.224 in $\mathbf{MCE}_{1.2}$ and -0.151 in $\mathbf{MCE}_{1.3}$.

5.1.1 Implications of $\mathbf{MCE}_1$

The main finding from the first class of MCEs is that the Heckman method seems to have the best performance for these relatively simple DGPs. While the Wooldridge method in most cases has higher ABs on APE(ρ) compared to the Heckman method the MAE and RMSEs are very similar between the models. It is clear that the LPM needs $T > 3$ to be able to consistently estimate the APE. The LPM records the highest observed AB at 516.16% and with poor performance for $\mathbf{MCE}_{1.3}$ where x_{it} follows a Nerlove process. The large ABs seems to be a result of that many of the response probabilities ($\boldsymbol{\beta x}_{it} + \rho y_{it} + c_i$) are outside the unit interval. Interestingly the ratio of observations outside the unit interval in $\mathbf{MCE}_{1.3}$ is only 1.2 times that of $\mathbf{MCE}_{1.1}$ and $\mathbf{MCE}_{1.2}$. This indicates that there is not only the share of response probabilities outside the unit interval that determines the bias but also how far away from the unit interval they are.

5.2 Finite sample results based on $\mathbf{MCE}_2$

In table 3 I have also included the results from $\mathbf{MCE}_{1.1}$ in the first panel. By comparing $\mathbf{MCE}_{2.1}$ through $\mathbf{MCE}_{2.3}$ with $\mathbf{MCE}_{1.1}$ we see that the only difference in specifications is that the second class of MCEs have non-normal error distributions. By comparing the finite sample properties of these different specifications we can thus shed light on how the three methods fare when the true DGP does not lead to a response probability determined by the standard normal CDF. As mentioned for the four preceding MCEs the ABs, MAEs and RMSEs of the LPM drops significantly for $T > 3$ when estimating APE(ρ). In the following I therefore ignore the LPM estimates for $T = 3$ and conclude that model does not perform satisfactory for $T = 3$. To make comparisons easier after each range I include the relevant range from $\mathbf{MCE}_{1.1}$ in parentheses.

For $\mathbf{MCE}_{2.1}$, where u_{it} follows a Laplace distribution, we see that the ABs of the estimated APE(ρ) range from 16.49% to 19.51% (1.59 to 5.36%) for the LPM, -6.52% to 2.62% (-0.83% to 3.53%) for the Heckman method and 17.24% to 21.97% (5.26% to 9.50%) for the Wooldridge method. All methods have relatively large increases in the ABs, but the Heckman method still performs well with an averaged AB just under 5%. APE(β_1) is relatively precisely estimated by the LPM with ABs ranging from 5.97% to 6.76% (-0.35% to 1.24%) compared to ABs ranging from -11.39% to 2.14% (-0,52% to 5.57%) while the ABs range from 14.08 to 16.59% (7.95 to 9.25%) for the Wooldridge method. The Heckman method has the lowest ABs for both APE(ρ) and APE(β_1). The MAE and RMSEs are lowest for the Heckman method for all T, followed by the LPM while the Wooldridge method has the highest error measurements. All three estimation methods have weaker performance than in $\mathbf{MCE}_{1.1}$.

For $\mathbf{MCE}_{2.2}$, where u_{it} follows a Laplace distribution, the results are more similar between the methods and comparable to those of $\mathbf{MCE}_{1.1}$. The ABs of the estimated APE(ρ) range from -3.21% to 9.03% (1.59% to 5.36%) for the LPM, from -3.71% to 5.74% (-0.83% to 3.53%) for the Heckman method and from 4.58% to 11.69% (5.26% to 9.50%) for the Wooldridge method. We thus see a slight worsening of performance as measured by the ABs compared to $\mathbf{MCE}_{1.1}$. APE(β_1) is again precisely estimated, with ABs ranging from only -0.58% to 0.38% (-0.35% to 1.24%) for the LPM, while the Heckman method has ABs ranging from 0.48 to 6.99% (-0.52 to 5.57%) and the Wooldridge method has ABs ranging from 7.16% to 9.76% (7.95% to 9.25%). Interestingly we see that the LPM and the Heckman method have roughly equal performance when estimating state dependence, while the Wooldridge method performs noticeably worse. The LPM is the most precise estimator for APE(β_1), while the ABs of the Wooldridge and Heckman methods is increasing in T. The RMSEs and MAEs are decreasing in T for all three methods.

For $\mathbf{MCE}_{2.3}$, where u_{it} follows a logistic distribution, there are some changes in performance compared to the earlier MCEs. The ABs of the estimated APE(ρ) are now ranging from -7.57% to 3.04% (1.59% to 5.36%) for the LPM, from 5.97% to 9.92% (-0.83% to 3.53%) for the Heckman method and from 3.57% to 6.77% (5.26% to 9.50%) for the Wooldridge method. Thus the performance for the Wooldridge improves while it deteriorates for the Heckman method. The LPM now underestimates the AB for $T > 4$. Again the LPM estimates APE(β_1) very precisely, with ABs ranging from -1.53% to 0.24% (-0.35% to 1.24%), while the Heckman method and the Wooldridge method have ABs ranging from 5.23% to 9.34% (-0,52% to 5.57%) and 4.04% to 6.25% (7.95% to 9.25%), respectively. Thus the Wooldridge method also estimates the APEs of x_{it} more precisely in $\mathbf{MCE}_{2.3}$, where the error term is misspecified, than in $\mathbf{MCE}_{1.1}$. The MAEs and RMSEs of the LPM and Wooldridge method decrease in T, for both APE estimates. For the Heckman method the MAEs and RMSEs of $\widehat{\text{APE}}(\rho)$ decrease, while they stay constant for $\widehat{\text{APE}}(\beta_1)$.

Table 3: Finite sample APE results for $\mathbf{MCE}_2$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Linear probability model						Heckman						Wooldridge 4 ($\mathbf{W}_4$)					
	ρ			β_1			ρ			β_1			ρ			β_1		
T	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE
	$\mathbf{MCE}_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$																
3	23.01	0.15	0.19	1.24	0.02	0.03	-2.21	0.06	0.07	-0.52	0.03	0.03	8.44	0.07	0.09	8.23	0.03	0.03
4	1.59	0.08	0.10	0.07	0.02	0.02	1.76	0.04	0.05	2.50	0.02	0.02	9.50	0.05	0.06	7.95	0.02	0.03
5	5.36	0.06	0.07	-0.35	0.01	0.02	0.84	0.04	0.05	3.62	0.02	0.02	7.28	0.04	0.05	8.28	0.02	0.02
8	4.41	0.03	0.04	0.14	0.01	0.01	0.22	0.02	0.03	4.16	0.01	0.02	7.37	0.02	0.03	9.14	0.02	0.02
12	1.65	0.02	0.03	-0.05	0.01	0.01	-0.83	0.02	0.02	4.54	0.01	0.01	5.26	0.02	0.02	9.25	0.02	0.02
20	4.57	0.02	0.02	-0.05	0.01	0.01	3.53	0.01	0.02	5.57	0.01	0.01	7.69	0.01	0.02	9.20	0.02	0.02
	$\mathbf{MCE}_{2.1}$	$u_{it} \sim \text{Laplace}\left(0, \sqrt{\frac{1}{2}}\right)$																
3	33.22	0.15	0.20	6.76	0.03	0.04	-6.09	0.06	0.07	-11.39	0.04	0.06	21.97	0.07	0.09	14.08	0.03	0.04
4	16.52	0.08	0.10	5.97	0.02	0.03	-6.52	0.04	0.05	-7.46	0.03	0.04	19.10	0.04	0.06	14.95	0.03	0.04
5	21.88	0.06	0.08	6.43	0.02	0.02	-4.41	0.03	0.04	-5.62	0.02	0.03	18.87	0.04	0.05	15.04	0.03	0.04
8	16.49	0.03	0.04	6.24	0.01	0.02	-6.60	0.02	0.03	-5.25	0.02	0.02	17.81	0.03	0.03	15.92	0.03	0.04
12	17.31	0.03	0.03	6.51	0.01	0.02	-3.99	0.02	0.02	-2.89	0.01	0.01	17.25	0.02	0.03	16.33	0.04	0.04
20	19.51	0.02	0.03	6.81	0.02	0.02	2.62	0.01	0.01	2.14	0.01	0.01	17.24	0.02	0.02	16.59	0.04	0.04
	$\mathbf{MCE}_{2.2}$	$u_{it} \sim \text{U}\left(\frac{-\sqrt{12}}{2}, \frac{\sqrt{12}}{2}\right)$																
3	18.80	0.14	0.18	0.05	0.03	0.03	-3.71	0.06	0.07	0.48	0.03	0.03	5.20	0.07	0.09	7.16	0.02	0.03
4	9.03	0.07	0.09	0.17	0.02	0.02	5.74	0.04	0.06	4.91	0.02	0.02	11.69	0.05	0.06	8.45	0.02	0.02
5	-3.21	0.05	0.07	0.38	0.01	0.02	-0.79	0.04	0.05	4.89	0.02	0.02	5.32	0.04	0.05	9.33	0.02	0.02
8	-0.64	0.03	0.04	-0.58	0.01	0.01	-0.67	0.02	0.03	5.26	0.01	0.02	4.58	0.02	0.03	8.93	0.02	0.02
12	0.15	0.02	0.03	-0.05	0.01	0.01	-0.13	0.02	0.02	5.84	0.01	0.02	4.95	0.02	0.02	9.53	0.02	0.02
20	1.97	0.01	0.02	0.25	0.01	0.01	2.51	0.01	0.02	6.99	0.01	0.02	5.66	0.01	0.02	9.76	0.02	0.02
	$\mathbf{MCE}_{2.3}$	$u_{it} \sim \text{Log}(0,1)$																
3	19.36	0.12	0.16	-0.14	0.03	0.03	-0.99	0.07	0.08	5.23	0.02	0.02	-0.96	0.07	0.09	4.04	0.02	0.03
4	3.04	0.07	0.09	0.03	0.02	0.02	5.97	0.05	0.06	7.61	0.02	0.02	5.77	0.05	0.06	5.52	0.02	0.02
5	-4.92	0.05	0.07	-1.53	0.02	0.02	9.92	0.04	0.05	7.56	0.02	0.02	6.74	0.04	0.05	4.38	0.02	0.02
8	-7.57	0.03	0.04	-0.18	0.01	0.01	6.88	0.02	0.03	9.09	0.02	0.02	3.81	0.02	0.03	5.62	0.01	0.01
12	-4.18	0.02	0.03	0.24	0.01	0.01	7.73	0.02	0.02	9.34	0.02	0.02	5.32	0.02	0.02	6.25	0.01	0.01
20	-4.53	0.02	0.02	0.06	0.01	0.01	6.19	0.01	0.02	9.18	0.02	0.02	3.57	0.01	0.02	6.07	0.01	0.01

For detailed description of each MCE see section 4.2. $\mathbf{W}_4$ converged in each instance. Heckmans method failed to converge in 15 instances for $\mathbf{MCE}_{1.1}$, 36 instances for $\mathbf{MCE}_{2.1}$, 15 instances for $\mathbf{MCE}_{2.2}$ and 27 instances for $\mathbf{MCE}_{2.3}$. The response probability outside the unit interval in 74.7% of repetitions in $\mathbf{MCE}_{1.1}$, in 74.9% of repetitions in $\mathbf{MCE}_{2.1}$, in 74.7% of repetitions in $\mathbf{MCE}_{2.2}$ and in 74.3% of repetitions in $\mathbf{MCE}_{2.3}$.

5.2.1 Implications of $\mathbf{MCE}_2$

In $\mathbf{MCE}_2$ I have analyzed the properties of the estimators when the error term is non-normally distributed. The main finding is that this violation of a necessary assumption in both the Wooldridge and Heckman models has only a small impact on the estimated APEs, just as in static models. Interestingly the LPM was also affected by the change in distributions, even though consistency or unbiasedness of the LPM does not place restrictions on the distribution of the error term. Even so we saw that the LPM had large biases, around 18% when the idiosyncratic error followed a Laplace distribution. On the other hand when the error term followed a uniform distribution the LPM performed better than in any of the other MCEs, thus error term distributions also affects the LPMs performance, a non-obvious result considering that each MCE has approximately the same number of observations where the response probability is in the unit interval. The second main finding is that the Heckman method also has the best finite sample performance when we experiment with the error term distribution, and it was in general less affected than the Wooldridge method, for unclear reasons. A third finding, in line with the findings from $\mathbf{MCE}_1$ is that the MAEs and RMSEs generally decline with T even though the ABs do not always decline.

5.3 Finite sample results based on $\mathbf{MCE}_3$

In table 4 I have again included the results from $\mathbf{MCE}_{1.1}$ in the first panel. For each MCE in $\mathbf{MCE}_3$ there is only one difference in the specifications compared to $\mathbf{MCE}_{1.1}$. We can thus isolate the effects of non-normal distributions of c_i, endogenous covariates and omitted relevant variables by comparing the results to those of $\mathbf{MCE}_{1.1}$. I again ignore the LPM results for $T = 3$. The relevant results from $\mathbf{MCE}_{1.1}$ are included in parentheses. u_{it} follows a standard normal distribution in all of the specifications in $\mathbf{MCE}_3$.

In $\mathbf{MCE}_{3.1}$ the individual-specific fixed effect c_i follows a standardized Gamma distribution. The LPM does not place any restrictions on c_i, while the Wooldridge and Heckman method both assume a standard normal distribution, although they both can be modified to allow for other distributions. The ABs of the estimated APE(ρ) range from 2.56% to 9.84% (1.59% to 5.36%) for the LPM, from 0.56% to 4.00% (-0.83% to 3.53%) for the Heckman method and from 6.64% to 10.23% (5.26% to 9.50%) for the Wooldridge method. The ABs increased slightly for all three methods. The ABs of APE(β_1) are almost unaffected, ranging from -0.45% to 1.13% (-0.35% to 1.24%) for the LPM, 1.86% to

6.22% (-0,52% to 5.57%) for the Heckman method and 7.42% to 8.80% (7.95 to 9.25%) for the Wooldridge method. The MAEs and RMSEs are smallest for the Heckman method for $T \leq 5$. When $T > 5$ the MAEs and RMSEs for the models are very similar and decrease with T for all three estimators.

In $\mathbf{MCE}_{3.2}$ x_{it} is endogenous, having a correlation coefficient of 0.315 with c_i. From section 2.4 we know that this should not affect the LPM but the effects on the Heckman and Wooldridge methods are less clear. The ABs of the estimated APE(ρ) range from 9.47% to 14.46% (1.59% to 5.36%) for the LPM, from 29.43 to 148.44% (-0.83% to 3.53%) for the Heckman method and from 1.03% to 10.18% (5.26% to 9.50%). The ABs increase slightly for the LPM, increase drastically for the Heckman method and decrease for the Wooldridge method. The ABs of APE(β_1) display a similar pattern, ranging from -0.31% to 1.53% (-0.35% to 1.24%) for the LPM, 24.41% to 30.07% (-0,52% to 5.57%) for the Heckman method and 2.40% to 5.11% (7.95 to 9.25%) for the Wooldridge method. The ABs are virtually unchanged for the LPM while they increased for the Heckman method and decreased for the Wooldridge method. These changes in ABs are also reflected in the error measurements. For APE(ρ) the Wooldridge method now has the lowest MAEs and RMSEs in every period and Heckman the highest. For APE(β_1) the Wooldridge method and LPM are tied for the lowest scores. Both the Heckman and Wooldridge methods show increasing performance with T.

In $\mathbf{MCE}_{3.3}$ the true DGP includes both $x_{1,it}$ and $x_{2,it}$, which are multivariate normally distributed with a correlation coefficient of 0.3. Note that $x_{2,it}$ and $y_{i,t-1}$ are not correlated. By omitting $x_{2,it}$ in the regressions I thus create an omitted variable problem. The ABs of the estimated APE(ρ) range from -4.19% to 2.11% (1.59% to 5.36%) for the LPM, from 0.88 to 8.20% (-0.83% to 3.53%) for the Heckman method and from 0.94% to 8.51% (5.26% to 9.50%). Thus the estimated APE(β_1) is largely unaffected by the omitted variable. The ABs of APE(β_1) on the other hand are affected with ABs ranging from 29.01% to 31.40% (-0.35% to 1.24%) for the LPM, 35.37% to 39.62% (-0,52% to 5.57%) for the Heckman method and 36.51% to 37.58% (7.95% to 9.25%) for the Wooldridge method. The LPM display the lowest ABs while the ABs of the Heckman and Wooldridge methods overlap. The MAEs and RMSEs for the estimated APE(ρ) are lowest for the Heckman method. For APE(β_1) the LPM has marginally lower MAEs and RMSEs than the two other methods. $\mathbf{MCE}_{3.3}$ present some comforting results as the state dependence is still precisely estimated in all models even when there are omitted variables that are correlated with the other explanatory variable.

Table 4: Finite sample APE results for $\mathbf{MCE}_3$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Linear probability model						Heckman						Wooldridge 4 ($\mathbf{W}_4$)					
	ρ			β_1			ρ			β_1			ρ			β_1		
T	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE	Average Bias %	MAE	RMSE
	$\mathbf{MCE}_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$																
3	23.01	0.15	0.19	1.24	0.02	0.03	-2.21	0.06	0.07	-0.52	0.03	0.03	8.44	0.07	0.09	8.23	0.03	0.03
4	1.59	0.08	0.10	0.07	0.02	0.02	1.76	0.04	0.05	2.50	0.02	0.02	9.50	0.05	0.06	7.95	0.02	0.03
5	5.36	0.06	0.07	-0.35	0.01	0.02	0.84	0.04	0.05	3.62	0.02	0.02	7.28	0.04	0.05	8.28	0.02	0.02
8	4.41	0.03	0.04	0.14	0.01	0.01	0.22	0.02	0.03	4.16	0.01	0.02	7.37	0.02	0.03	9.14	0.02	0.02
12	1.65	0.02	0.03	-0.05	0.01	0.01	-0.83	0.02	0.02	4.54	0.01	0.01	5.26	0.02	0.02	9.25	0.02	0.02
20	4.57	0.02	0.02	-0.05	0.01	0.01	3.53	0.01	0.02	5.57	0.01	0.01	7.69	0.01	0.02	9.20	0.02	0.02
	$\mathbf{MCE}_{3.1}$	$c_i \sim \Gamma(4, 0.5) - 2$																
3	22.37	0.13	0.18	1.13	0.03	0.03	0.56	0.06	0.07	1.86	0.02	0.03	10.23	0.07	0.09	7.49	0.03	0.03
4	9.84	0.08	0.10	0.55	0.02	0.02	3.90	0.04	0.06	4.35	0.02	0.02	9.00	0.05	0.06	7.67	0.02	0.03
5	5.80	0.06	0.07	-0.45	0.01	0.02	2.84	0.03	0.04	4.92	0.02	0.02	7.40	0.04	0.05	7.42	0.02	0.02
8	4.78	0.03	0.04	0.83	0.01	0.01	2.30	0.02	0.03	6.06	0.02	0.02	7.02	0.02	0.03	8.80	0.02	0.02
12	2.56	0.02	0.03	-0.01	0.01	0.01	2.52	0.02	0.02	5.70	0.01	0.02	6.64	0.02	0.02	8.34	0.02	0.02
20	4.69	0.01	0.02	0.02	0.01	0.01	4.00	0.01	0.02	6.22	0.01	0.02	6.76	0.01	0.02	8.40	0.02	0.02
	$\mathbf{MCE}_{3.2}$	$x_{it} \sim \mathcal{N}\left(\frac{c_i}{3}, 1\right)$																
3	37.45	0.15	0.20	1.53	0.03	0.03	148.44	0.15	0.16	28.01	0.05	0.06	10.18	0.07	0.09	5.11	0.02	0.03
4	12.72	0.09	0.11	0.10	0.02	0.02	124.13	0.13	0.14	29.36	0.06	0.06	6.98	0.05	0.06	4.71	0.01	0.02
5	14.46	0.06	0.07	0.46	0.01	0.02	101.36	0.10	0.11	30.07	0.06	0.06	3.91	0.03	0.04	4.83	0.01	0.02
8	9.47	0.03	0.04	-0.31	0.01	0.01	60.98	0.06	0.07	27.20	0.05	0.05	2.68	0.02	0.03	3.87	0.01	0.01
12	10.28	0.02	0.03	-0.22	0.01	0.01	41.80	0.04	0.05	24.41	0.05	0.05	2.09	0.02	0.02	3.11	0.01	0.01
20	11.40	0.02	0.02	0.18	0.01	0.01	29.43	0.03	0.03	22.04	0.04	0.04	1.03	0.01	0.01	2.40	0.01	0.01
	$\mathbf{MCE}_{3.3}$	$x_{1,it} \sim \mathcal{N}(0,1)$, $x_{2,it} \sim \mathcal{N}(0,1)$, $\text{Corr}(x_{1,it}, x_{2,it}) = 0.3$																
3	24.37	0.12	0.16	31.40	0.06	0.07	0.88	0.06	0.07	35.37	0.06	0.07	0.94	0.07	0.09	36.95	0.07	0.07
4	0.03	0.07	0.09	29.95	0.05	0.06	5.58	0.04	0.05	37.81	0.07	0.07	5.86	0.04	0.06	37.05	0.07	0.07
5	2.11	0.05	0.06	29.07	0.05	0.06	8.20	0.03	0.04	38.21	0.07	0.07	8.51	0.03	0.04	36.51	0.07	0.07
8	-4.19	0.03	0.04	29.81	0.05	0.06	5.56	0.02	0.03	39.75	0.07	0.07	4.31	0.02	0.03	37.55	0.07	0.07
12	-0.48	0.02	0.03	29.87	0.05	0.06	5.87	0.02	0.02	39.64	0.07	0.07	5.00	0.02	0.02	37.55	0.07	0.07
20	-1.96	0.01	0.02	29.99	0.05	0.05	4.98	0.01	0.02	39.62	0.07	0.07	3.83	0.01	0.01	37.58	0.07	0.07

For detailed description of each MCE see section 4.3. $\mathbf{W}_4$ converged in each instance. Heckmans method failed to converge in 15 instances for $\mathbf{MCE}_{1.1}$, 20 instances for $\mathbf{MCE}_{3.1}$, 99 instances for $\mathbf{MCE}_{3.2}$ and 20 instances for $\mathbf{MCE}_{3.3}$. The response probability outside the unit interval in 74.7% of repetitions in $\mathbf{MCE}_{1.1}$, in 74.9% of repetitions in $\mathbf{MCE}_{3.1}$, in 78.6% of repetitions in $\mathbf{MCE}_{3.2}$ and in 80.0% of repetitions in $\mathbf{MCE}_{3.3}$.

5.3.1 Implications of $\mathbf{MCE}_3$

In $\mathbf{MCE}_3$ I have designed DGPs that shed preliminary light on some issues that are relevant for empirical work. Both the Wooldridge and Heckman methods require that one integrates out the individual-specific fixed effect. To do so one must specify its distribution, which is commonly assumed to be the standard normal. When c_i was Gamma distributed we saw that performance of all estimators decreased somewhat. Interestingly the LPM had the biggest reduction in performance when estimating state dependence. While more research is needed it seems that both the Wooldridge and Heckman methods are relatively robust to misspecification of the distribution of c_i. In $\mathbf{MCE}_{3.2}$ x_{it} was endogenous with a positive correlation with c_i. For some reason this also affected the performance of the LPM somewhat, but drastically affected the Heckman method which then overestimated the state dependence, with ABs between 30% and 150%. The Wooldridge method was largely unaffected by this endogeneity, as Rabe-Hesketh and Skrondal (2013) showed. Endogeneity is thus a serious concern in the Heckman method. In the last experiment, $\mathbf{MCE}_{3.3}$ the DGP included two exogenous variables, where one was omitted in the regressions. The estimated APE(ρ)s were almost unaffected by this change, but the omitted variable led to significant decreases when estimating APE(β_1).

5.4 Coefficients results

As discussed the LPM cannot identify the coefficients of the latent variable while both the Wooldridge and Heckman method do so. In the earlier literature these methods have been evaluated on their performance when estimating coefficients instead of APEs. In tables 7, 8 and 9 in appendix C I have reported the finite sample results on the coefficient estimates for the Heckman and Wooldridge methods. Below I provide a short discussion on the finite sample performance of the Heckman and Wooldridge ($\mathbf{W}_4$) methods when estimating the coefficients of the latent variable.

The Heckman method and $\mathbf{W}_4$ both have very similar performance on the coefficient estimates and in most cases their coefficient estimates do not differ noticeably. The only case where their estimated coefficients do differ significantly is for $\mathbf{MCE}_{3.2}$ where x_{it} is correlated with c_i. As we saw in section 5.3 this is also the only process where the estimated APEs of the Heckman method have large ABs. For $\mathbf{MCE}_{1.1}$,$\mathbf{MCE}_{1.2}$, $\mathbf{MCE}_{1.3}$, $\mathbf{MCE}_{2.2}$ and $\mathbf{MCE}_{3.1}$ the estimated coefficients have very low ABs, under 5%.

The ABs of the coefficients are around 50% when u_{it} is Laplace distributed with unity variance. Interestingly the ABs of the APEs were unbiased for both methods. When u_{it}

follows a logistic distribution both methods underestimate the coefficients with approximately 40%. By multiplying with the standard deviation of the logistic distribution (see appendix A) we see that the estimates are unbiased. For $\mathbf{MCE}_{3.2}$ the Heckman method overestimates ρ while the Wooldridge method is unbiased, which explains the discrepancy between the model's APEs for this MCE. Finally when there is an omitted variable that is correlated with $x_{1,it}$ but uncorrelated with $y_{i,t-1}$ both methods underestimate ρ with roughly 30%. The MAEs and RMSEs are comparable for the Heckman and Wooldridge methods in most MCEs.

6 Summary and conclusions

This thesis has evaluated three estimation methods that can be used to estimate state dependence. To evaluate the finite sample properties of these estimators nine different DGPs have been employed. Each DGP is designed in a manner that allows us to draw conclusions on the performance of the estimators in empirical settings. This is done in a method inspired by controlled experiments. By changing only one parameter or variable at a time I isolate the effect of the changes. The first experiment, $\mathbf{MCE}_{1.1}$, provides a control group. To for example identify the effect of a non-normal distribution on performance I compare the results from a DGP where the error term is non-normally distributed to the results from $\mathbf{MCE}_{1.1}$. By repeating the experiments 500 times we can then, with good confidence, say that any differences are due to the non-normal distribution.

The main finding is positive in that all methods perform relatively well on most DGPs. The Heckman method is usually the method with the best finite sample performance, with the lowest ABs, MAEs and RMSEs on the estimated APE(ρ). For $T > 3$ the LPM usually has slightly better estimation results than the Wooldridge method. While the Heckman method has the best overall finite sample performance its performance broke down when x_{it} was endogenous. This is a serious concern as we expect c_i to be correlated with the explanatory variables in empirical settings. While the LPM is largely unaffected by the endogeneity (as c_i is first-differenced away) its performance broke down when the DGP included a Nerlove process. This is another serious concern, as the Nerlove process is created to imitate the age trended variable often encountered in microeconomic data, on which these methods are often applied. The Wooldridge method was relatively robust to all the different DGPs with ABs, with a maximum AB of 21.97%, and most ABs are less than 10%.

There are some minor findings. To use the LPM one should have $T \geq 4$, even while consistency only requires three periods, as the AB drops significantly once there are four observed periods. For all methods the precision increases, as measured by the MAEs and RMSEs, when T increases, even when the ABs do not decrease. Furthermore there was a clear tendency for the methods to slightly overestimate the state dependence. Finally, even in cases where the Wooldridge and Heckman methods are biased for β_1 or ρ the ABs of the APEs are relatively small.

Based on these results I conclude that the Wooldridge method is the most robust method to different DGPs and violated assumptions. At the same time the Heckman method is usually the most precise estimator with the LPM coming second. When the LPM or the Heckman method broke down, the other two methods still performed well. Therefore to estimate the state dependence one should use all three estimation methods. If all three estimators provide similar results one should put most weight on the Heckman estimates. If there are large discrepancies and these are caused by endogeneity one should use the Wooldridge and the LPM results over those from the Heckman method. If the discrepancies are caused by age trended variables, so that the LPM estimates are heavily biased, one should use the Wooldridge and Heckman results over those from the LPM.

References

Akay, A. (2012) "Finite-sample comparison of alternative methods for estimating dynamic panel data models," *Journal of Applied Econometrics*, Vol. 27, pp. 1189–1204.

Anderson, T. W. and C. Hsiao (1981) "Estimation of dynamic models with error components," *Journal of the American statistical Association*, Vol. 76, pp. 598–606.

——— (1982) "Formulation and estimation of dynamic models using panel data," *Journal of econometrics*, Vol. 18, pp. 67–82.

Arellano, M. and S. Bond (1991) "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, Vol. 58, pp. 277–297.

Arulampalam, W. and M. Stewart (2009) "Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators*," *Oxford bulletin of economics and statistics*, Vol. 71, pp. 659–681.

Behr, A. (2003) "A comparison of dynamic panel data estimators: Monte Carlo evidence and an application to the investment function," Discussion paper 05/03, Economic Research Centre of the Deutsche Bundesbank.

Bhuller, M. and E. E. Brandsås (2013) "Fattigdomsdynamikk blant innvandrere (in Norwegian, Poverty dynamics among immigrants)," *SSB Rapporter*, Vol. 40/2013, pp. 1–59.

Biewen, M. (2009) "Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition," *Journal of Applied Econometrics*, Vol. 24, pp. 1095–1116.

Blundell, R. and S. Bond (1998) "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, Vol. 87, pp. 115–143.

Cameron, A. C. and P. K. Trivedi (2005) *Microeconometrics: methods and applications*: Cambridge university press, New York, NY.

Carro, J. M. (2007) "Estimating dynamic panel data discrete choice models with fixed effects," *Journal of Econometrics*, Vol. 140, pp. 503–528.

Chamberlain, G. (1980) "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, Vol. 47, pp. 225–238.

Chay, K. Y., H. W. Hoynes, and D. Hyslop (1999) "A non-experimental analysis of true state dependence in monthly welfare participation sequences," in *American Statistical Association, Proceedings of the Business and Economic Statistics Section*, pp. 9–17.

Fernández-Val, I. and M. Weidner (2013) "Individual and time effects in nonlinear panel models with large N, T," CeMMAP working paper series.

Heckman, J. J. (1981a) "Heterogeneity and state dependence," in S. Rosen ed. *Studies in labor markets*: University of Chicago Press, Chicago, IL, pp. 91–140.

——— (1981b) "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence," in C.F. Manski and D. McFadden eds. *Structural Analysis of Discrete Data with Econometric Applications*: MIT press, Cambridge, MA, pp. 179–195.

Honoré, B. E. and E. Kyriazidou (2000) "Panel data discrete choice models with lagged dependent variables," *Econometrica*, Vol. 68, pp. 839–874.

Horrace, W. C. and R. L. Oaxaca (2006) "Results on the bias and inconsistency of ordinary least squares for the linear probability model," *Economics Letters*, Vol. 90, pp. 321–327.

Hyslop, D. R. (1999) "State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women," *Econometrica*, Vol. 67, pp. 1255–1294.

Neyman, J. and E. L. Scott (1948) "Consistent estimates based on partially consistent observations," *Econometrica: Journal of the Econometric Society*, Vol. 16, pp. 1–32.

Rabe-Hesketh, S. and A. Skrondal (2013) "Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem," *Economics Letters*, Vol. 120, pp. 346–349.

Ruud, P. A. (1983) "Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models," *Econometrica: Journal of the Econometric Society*, Vol. 51, pp. 225–228.

Stewart, M. (2006) "-redprob-: A Stata program for the Heckman estimator of the random effects dynamic probit model," Mimeo, University of Warwick.

——— (2007) "The interrelated dynamics of unemployment and low-wage employment," *Journal of Applied Econometrics*, Vol. 22, pp. 511–531.

Wooldridge, J. M. (2005) "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of applied econometrics*, Vol. 20, pp. 39–54.

——— (2010) *Econometric analysis of cross section and panel data*: MIT press, Cambridge, MA, 2nd edition.

Appendices

A On normalizations in binary choice models

Above I assumed that the both the distribution and the variance of the idiosyncratic error is known. In most cases the idiosyncratic error is assumed to be either standard normally distributed and thus with unity variance or logistically distributed with variance

$\pi^2/3$. The normalizations on the variance of the idiosyncratic unfortunately prohibits identification of both the coefficients and the variance: Let ε be the idiosyncratic error, which is normally distributed with variance C^2. I simplify the notation so that $y_{it}^* = \boldsymbol{\beta}\boldsymbol{x}_{it} + \varepsilon_{it}$. Following the steps above we get

$$\mathrm{P}(y_{it} = 1) = \cdots = P(\varepsilon > -\boldsymbol{\beta}\boldsymbol{x}_{it}) = G(\boldsymbol{\beta}\boldsymbol{x}_{it}) \tag{48}$$

But if we want to employ the probit model we know that G must be the CDF of standard normal distribution, Φ. If we multiply the latent variable with $\frac{1}{C}$, we get $\frac{y_{it}^*}{C} = \frac{\boldsymbol{\beta}\boldsymbol{x}_{it}}{C} + \frac{\varepsilon_{it}}{C}$. Importantly this is the same model with unchanged variables, the only change being that the coefficients are scaled differently. Then, letting $u_{it} = \frac{\varepsilon}{C}$ we see that u_{it} is a standard normally distributed variable. The constant scaling factor C does not affect the distribution or mean of u_{it}, and the variance of u_{it} is:

$$\mathrm{var}(u_{it}) = \mathrm{var}\left(\frac{\varepsilon}{C}\right) = \frac{\mathrm{Var}(\varepsilon)}{C^2} = \frac{C^2}{C^2} = 1, \tag{49}$$

which gives:

$$\mathrm{P}(y_{it} = 1) = \cdots = P\left(u_{it} > -\frac{\boldsymbol{\beta}}{C}\boldsymbol{x}_{it}\right) = G\left(\frac{\boldsymbol{\beta}}{C}\boldsymbol{x}_{it}\right) = \Phi\left(\frac{\boldsymbol{\beta}}{C}\boldsymbol{x}_{it}\right) \tag{50}$$

This means that the probit estimates do *not* estimate the coefficients of the latent variable, but rather the coefficients divided by the standard deviation of the idiosyncratic error term. Thus, we only consistently estimate the coefficients if the idiosyncratic variance is unity.

On the other hand the normalization on the intercept does not affect the coefficients of the structural variables. In section 2 the threshold value for y_{it} was set to zero, i.e. $y_{it} = 1\{y_{it}^* > 0\}$. If we let the threshold value be a constant C instead:

$$y_{it} = 1\{y_{it}^* > C\} \tag{51}$$

To simplify notation I let $y_{it}^* = \beta_0 + \boldsymbol{\beta}\boldsymbol{x}_{it} + u_{it}$, where I have extracted the constant term from the vectors. We then get

$$\mathrm{P}(y_{it} = 1) = \mathrm{P}(y_{it}^* > C) = P(\beta_0 + \boldsymbol{\beta}\boldsymbol{x}_{it} + u_{it} > C) \tag{52}$$

Collecting the terms we get:

$$= \mathrm{P}([\beta_0 - C] + \boldsymbol{\beta}\boldsymbol{x}_{it} + u_{it} > 0) = \mathrm{P}(u_{it} > -[\beta_0 - C] - \boldsymbol{\beta}\boldsymbol{x}_{it}) \tag{53}$$

The (true) difference $\beta_0 - C$ is unknown, and as such C can be set arbitrarily as it only scales the intercept.

B APEs for the remaining Wooldridge methods

In tables 5 and 6 I have attached the finite sample results for $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$ and $\mathbf{W}_5$. By comparing the results with each other on sees clearly that there are only small differences between the different Wooldridge specifications. Even so, the $\mathbf{W}_2$ has lower ABs than $\mathbf{W}_1$ in every single MCE. Similarly $\mathbf{W}_4$ has lower bias than $\mathbf{W}_3$ in all MCEs. This is a clear indication of that we get efficiency gains by including x_{i1} in the reduced form for c_i. Finally $\mathbf{W}_5$ do not pose any clear efficiency gains compared to $\mathbf{W}_2$ and $\mathbf{W}_4$. The only MCE for which the $\mathbf{W}_5$ has a noticeably different performance is when the individual-specific fixed effect follows a gamma distribution, $\mathbf{MCE}_{3.1}$, where it performs worse than $\mathbf{W}_4$.

Table 5: Finite sample APE results for $\mathbf{W}_1$ and $\mathbf{W}_2$.

	Wooldridge 1						Wooldridge 2					
	ρ			β_1			ρ			β_1		
	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE
	$\mathbf{MCE}_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$										
3	10.87	0.08	0.10	9.25	0.03	0.03	9.82	0.08	0.10	8.40	0.03	0.03
4	10.09	0.05	0.06	9.03	0.02	0.03	9.60	0.05	0.06	8.06	0.02	0.03
5	7.88	0.04	0.05	9.31	0.02	0.02	7.16	0.04	0.05	8.26	0.02	0.02
8	7.90	0.02	0.03	9.82	0.02	0.02	7.26	0.02	0.03	8.85	0.02	0.02
12	5.59	0.02	0.02	9.67	0.02	0.02	4.88	0.02	0.02	8.67	0.02	0.02
20	7.60	0.01	0.02	9.06	0.02	0.02	6.93	0.01	0.02	8.11	0.02	0.02
	$\mathbf{MCE}_{1.2}$	$x_{it} \sim \frac{\chi_1^2 - 1}{\sqrt{2}}$										
3	14.54	0.09	0.11	9.53	0.04	0.05	13.14	0.09	0.11	8.43	0.04	0.04
4	12.99	0.06	0.07	11.95	0.03	0.04	12.37	0.06	0.07	10.88	0.03	0.04
5	9.75	0.04	0.05	11.08	0.03	0.03	9.23	0.04	0.05	9.99	0.03	0.03
8	9.02	0.03	0.03	11.72	0.03	0.03	8.39	0.03	0.03	10.68	0.03	0.03
12	8.99	0.02	0.03	11.61	0.03	0.03	8.29	0.02	0.03	10.57	0.02	0.03
20	8.69	0.02	0.02	11.94	0.03	0.03	8.00	0.02	0.02	10.94	0.03	0.03
	$\mathbf{MCE}_{1.3}$	$x_{it} = 0.5x_{i,t-1} + \frac{t+24}{10} + \text{U}(\frac{-1}{2}, \frac{1}{2})$										
3	15.66	0.10	0.12	1.44	0.07	0.09	14.73	0.10	0.12	1.26	0.07	0.09
4	16.81	0.06	0.08	2.82	0.04	0.05	16.78	0.06	0.08	2.71	0.04	0.05
5	9.81	0.04	0.06	4.88	0.03	0.04	9.82	0.04	0.06	4.70	0.03	0.04
8	0.84	0.02	0.03	2.44	0.02	0.02	0.86	0.02	0.03	2.43	0.01	0.02
12	-0.23	0.02	0.02	-0.91	0.01	0.01	-0.12	0.02	0.02	-0.81	0.01	0.01
20	-2.25	0.01	0.01	-2.36	0.00	0.01	-2.09	0.01	0.01	-2.22	0.00	0.01
	$\mathbf{MCE}_{2.1}$	$u_{it} \sim \text{Laplace}\left(0, \sqrt{\frac{1}{2}}\right)$										
3	25.16	0.08	0.10	16.12	0.04	0.05	22.91	0.07	0.09	14.10	0.03	0.04
4	20.94	0.04	0.06	17.05	0.04	0.04	19.16	0.04	0.06	15.02	0.03	0.04
5	20.60	0.04	0.05	17.13	0.04	0.04	18.81	0.04	0.05	15.01	0.03	0.04
8	19.06	0.03	0.03	17.59	0.04	0.04	17.66	0.03	0.03	15.56	0.03	0.04
12	18.32	0.02	0.03	17.68	0.04	0.04	16.79	0.02	0.03	15.62	0.03	0.04
20	17.74	0.02	0.02	17.20	0.04	0.04	16.38	0.02	0.02	15.35	0.03	0.03
	$\mathbf{MCE}_{2.2}$	$u_{it} \sim \text{U}\left(\frac{-\sqrt{12}}{2}, \frac{\sqrt{12}}{2}\right)$										
3	5.88	0.08	0.10	8.26	0.03	0.03	4.72	0.08	0.09	7.25	0.03	0.03
4	13.20	0.05	0.06	9.45	0.02	0.03	12.59	0.05	0.06	8.58	0.02	0.03
5	5.73	0.04	0.05	10.20	0.02	0.03	5.23	0.04	0.05	9.32	0.02	0.02
8	5.03	0.02	0.03	9.52	0.02	0.02	4.50	0.02	0.03	8.65	0.02	0.02
12	5.13	0.02	0.02	9.78	0.02	0.02	4.57	0.02	0.02	8.95	0.02	0.02
20	5.47	0.01	0.02	9.47	0.02	0.02	4.92	0.01	0.02	8.70	0.02	0.02
	$\mathbf{MCE}_{2.3}$	$u_{it} \sim \text{Log}(0,1)$										
3	-0.45	0.08	0.10	4.44	0.02	0.03	-1.75	0.08	0.09	4.17	0.02	0.03
4	6.55	0.05	0.06	5.91	0.02	0.02	6.24	0.05	0.06	5.62	0.02	0.02
5	6.87	0.04	0.05	4.62	0.02	0.02	6.71	0.04	0.05	4.36	0.02	0.02
8	3.79	0.02	0.03	5.64	0.01	0.01	3.62	0.02	0.03	5.41	0.01	0.01
12	5.17	0.02	0.02	6.03	0.01	0.01	5.02	0.02	0.02	5.82	0.01	0.01
20	3.20	0.01	0.02	5.50	0.01	0.01	3.04	0.01	0.02	5.30	0.01	0.01
	$\mathbf{MCE}_{3.1}$	$c_i \sim \Gamma(4, 0.5) - 2$										
3	13.03	0.08	0.10	8.38	0.03	0.03	11.81	0.08	0.10	7.58	0.03	0.03
4	10.58	0.05	0.06	8.65	0.02	0.03	9.36	0.05	0.06	7.75	0.02	0.03
5	8.40	0.04	0.05	8.28	0.02	0.02	7.85	0.04	0.05	7.38	0.02	0.02
8	7.51	0.02	0.03	9.43	0.02	0.02	6.97	0.02	0.03	8.51	0.02	0.02
12	6.86	0.02	0.02	8.57	0.02	0.02	6.30	0.02	0.02	7.77	0.02	0.02
20	6.57	0.01	0.02	8.14	0.02	0.02	6.03	0.01	0.02	7.37	0.02	0.02
	$\mathbf{MCE}_{3.2}$	$x_{it} \sim \mathcal{N}\left(\frac{c_i}{3}, 1\right)$										
3	9.60	0.07	0.09	5.16	0.02	0.03	9.50	0.07	0.09	5.07	0.02	0.03
4	7.18	0.05	0.06	4.91	0.02	0.02	7.09	0.05	0.06	4.80	0.02	0.02
5	4.07	0.03	0.04	4.92	0.01	0.02	3.98	0.03	0.04	4.85	0.01	0.02
8	2.46	0.02	0.03	3.75	0.01	0.01	2.45	0.02	0.03	3.71	0.01	0.01
12	2.02	0.02	0.02	2.90	0.01	0.01	2.01	0.02	0.02	2.87	0.01	0.01
20	0.83	0.01	0.01	2.09	0.01	0.01	0.82	0.01	0.01	2.06	0.01	0.01
	$\mathbf{MCE}_{3.3}$	$x_{1,it} \sim \mathcal{N}(0,1),\ x_{2,it} \sim \mathcal{N}(0,1),\ \text{Corr}(x_{1,it}, x_{2,it}) = 0.3$										
3	2.80	0.07	0.09	37.72	0.07	0.07	1.36	0.07	0.09	37.18	0.07	0.07
4	5.82	0.05	0.06	37.67	0.07	0.07	5.37	0.05	0.06	37.16	0.07	0.07
5	9.03	0.04	0.05	37.03	0.07	0.07	8.62	0.03	0.04	36.48	0.07	0.07
8	4.32	0.02	0.03	37.74	0.07	0.07	4.07	0.02	0.03	37.26	0.07	0.07
12	5.00	0.02	0.02	37.53	0.07	0.07	4.72	0.02	0.02	37.03	0.07	0.07
20	3.61	0.01	0.01	37.12	0.07	0.07	3.36	0.01	0.01	36.67	0.07	0.07

Table 6: Finite sample APE results for $\mathbf{W}_3$ and $\mathbf{W}_5$.

	Wooldridge 3						Wooldridge 5					
	ρ			β_1			ρ			β_1		
	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE
	MCE$_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$										
3	9.28	0.08	0.09	9.09	0.03	0.03	8.62	0.07	0.09	8.13	0.03	0.03
4	10.06	0.05	0.06	8.93	0.02	0.03	9.58	0.05	0.06	7.85	0.02	0.02
5	7.96	0.04	0.05	9.35	0.02	0.03	7.25	0.04	0.05	8.20	0.02	0.02
8	8.08	0.02	0.03	10.16	0.02	0.02	7.33	0.02	0.03	9.07	0.02	0.02
12	6.02	0.02	0.02	10.32	0.02	0.02	5.21	0.02	0.02	9.19	0.02	0.02
20	8.44	0.01	0.02	10.26	0.02	0.02	7.65	0.01	0.02	9.14	0.02	0.02
	MCE$_{1.2}$	$x_{it} \sim \frac{\chi_1^2-1}{\sqrt{2}}$										
3	13.70	0.09	0.11	9.48	0.04	0.05	14.30	0.09	0.11	9.18	0.04	0.05
4	12.71	0.06	0.07	11.91	0.03	0.04	12.38	0.06	0.07	10.86	0.03	0.04
5	9.91	0.04	0.05	11.18	0.03	0.03	9.39	0.04	0.05	10.03	0.03	0.03
8	9.14	0.03	0.03	12.18	0.03	0.03	8.46	0.03	0.03	11.04	0.03	0.03
12	9.54	0.02	0.03	12.47	0.03	0.03	8.72	0.02	0.03	11.26	0.03	0.03
20	9.71	0.02	0.02	13.46	0.03	0.03	8.91	0.02	0.02	12.30	0.03	0.03
	MCE$_{1.3}$	$x_{it} = 0.5x_{i,t-1} + \frac{t+24}{10} + \mathrm{U}(\frac{-1}{2}, \frac{1}{2})$										
3	13.67	0.10	0.12	1.11	0.07	0.09	15.96	0.10	0.12	1.37	0.07	0.09
4	15.40	0.06	0.08	2.12	0.04	0.05	16.89	0.06	0.08	2.62	0.04	0.05
5	8.92	0.04	0.06	4.10	0.03	0.04	9.93	0.04	0.06	4.64	0.03	0.04
8	0.48	0.02	0.03	1.72	0.01	0.02	0.89	0.02	0.03	2.51	0.01	0.02
12	-0.96	0.02	0.02	-2.08	0.01	0.01	-0.54	0.02	0.02	-1.26	0.01	0.01
20	-3.67	0.01	0.01	-4.05	0.00	0.01	-3.19	0.01	0.01	-3.30	0.00	0.01
	MCE$_{2.1}$	$u_{it} \sim \mathrm{Laplace}\left(0, \sqrt{\frac{1}{2}}\right)$										
3	23.57	0.07	0.09	16.06	0.04	0.05	21.74	0.07	0.09	13.96	0.03	0.04
4	20.74	0.04	0.06	16.99	0.04	0.04	19.17	0.04	0.06	14.81	0.03	0.04
5	20.64	0.04	0.05	17.20	0.04	0.04	18.82	0.04	0.05	14.92	0.03	0.04
8	19.24	0.03	0.03	18.01	0.04	0.04	17.75	0.03	0.03	15.84	0.03	0.04
12	18.88	0.02	0.03	18.50	0.04	0.04	17.19	0.02	0.03	16.24	0.04	0.04
20	18.79	0.02	0.03	18.68	0.04	0.04	17.18	0.02	0.02	16.50	0.04	0.04
	MCE$_{2.2}$	$u_{it} \sim \mathrm{U}\left(\frac{-\sqrt{12}}{2}, \frac{\sqrt{12}}{2}\right)$										
3	6.57	0.07	0.09	8.16	0.03	0.03	4.95	0.07	0.09	7.00	0.02	0.03
4	12.32	0.05	0.06	9.33	0.02	0.03	11.60	0.05	0.06	8.31	0.02	0.02
5	5.82	0.04	0.05	10.22	0.02	0.03	5.31	0.04	0.05	9.25	0.02	0.02
8	5.17	0.02	0.03	9.84	0.02	0.02	4.53	0.02	0.03	8.85	0.02	0.02
12	5.56	0.02	0.02	10.43	0.02	0.02	4.90	0.02	0.02	9.47	0.02	0.02
20	6.28	0.01	0.02	10.63	0.02	0.02	5.61	0.01	0.02	9.69	0.02	0.02
	MCE$_{2.3}$	$u_{it} \sim \mathrm{Log}(0,1)$										
3	0.41	0.08	0.09	4.32	0.02	0.03	-1.03	0.07	0.09	3.89	0.02	0.03
4	6.21	0.05	0.06	5.81	0.02	0.02	5.64	0.05	0.06	5.43	0.02	0.02
5	6.90	0.04	0.05	4.64	0.02	0.02	6.69	0.04	0.05	4.29	0.02	0.02
8	3.99	0.02	0.03	5.86	0.01	0.02	3.74	0.02	0.03	5.55	0.01	0.01
12	5.47	0.02	0.02	6.48	0.01	0.01	5.29	0.02	0.02	6.19	0.01	0.01
20	3.73	0.01	0.02	6.30	0.01	0.01	3.53	0.01	0.02	6.02	0.01	0.01
	MCE$_{3.1}$	$c_i \sim \Gamma(4, 0.5) - 2$										
3	10.28	0.07	0.09	5.20	0.02	0.03	9.94	0.07	0.09	4.96	0.02	0.03
4	7.08	0.05	0.06	4.82	0.01	0.02	6.98	0.05	0.06	4.66	0.01	0.02
5	3.97	0.03	0.04	4.90	0.01	0.02	3.98	0.03	0.04	4.79	0.01	0.02
8	2.70	0.02	0.03	3.92	0.01	0.01	2.68	0.02	0.03	3.84	0.01	0.01
12	2.10	0.02	0.02	3.14	0.01	0.01	2.07	0.02	0.02	3.08	0.01	0.01
20	1.05	0.01	0.01	2.43	0.01	0.01	1.04	0.01	0.01	2.40	0.01	0.01
	MCE$_{3.2}$	$x_{it} \sim \mathcal{N}\left(\frac{c_i}{3}, 1\right)$										
3	10.28	0.07	0.09	5.20	0.02	0.03	9.94	0.07	0.09	4.96	0.02	0.03
4	7.08	0.05	0.06	4.82	0.01	0.02	6.98	0.05	0.06	4.66	0.01	0.02
5	3.97	0.03	0.04	4.90	0.01	0.02	3.98	0.03	0.04	4.79	0.01	0.02
8	2.70	0.02	0.03	3.92	0.01	0.01	2.68	0.02	0.03	3.84	0.01	0.01
12	2.10	0.02	0.02	3.14	0.01	0.01	2.07	0.02	0.02	3.08	0.01	0.01
20	1.05	0.01	0.01	2.43	0.01	0.01	1.04	0.01	0.01	2.40	0.01	0.01
	MCE$_{3.3}$	$x_{1,it} \sim \mathcal{N}(0,1),\ x_{2,it} \sim \mathcal{N}(0,1),\ \mathrm{Corr}(x_{1,it}, x_{2,it}) = 0.3$										
3	2.30	0.07	0.09	37.49	0.07	0.07	1.35	0.07	0.09	36.79	0.07	0.07
4	6.17	0.04	0.06	37.57	0.07	0.07	5.88	0.04	0.06	36.89	0.07	0.07
5	8.97	0.03	0.05	37.06	0.07	0.07	8.48	0.03	0.04	36.41	0.07	0.07
8	4.55	0.02	0.03	38.06	0.07	0.07	4.29	0.02	0.03	37.47	0.07	0.07
12	5.29	0.02	0.02	38.09	0.07	0.07	4.98	0.02	0.02	37.49	0.07	0.07
20	4.12	0.01	0.01	38.07	0.07	0.07	3.80	0.01	0.01	37.52	0.07	0.07

C Coefficient estimates for the Heckman and Wooldridge methods

In tables 7, 8 and 9 I have attached the finite sample results on the coefficient estimates for the Heckman and Wooldridge methods. The LPM does not estimate the coefficients of the latent variable only the APEs.

The Heckman method and $\mathbf{W}_4$ both have very similar performance on the coefficient estimates and in most cases their coefficient estimates do not differ noticeably. The only case where their estimated coefficients do differ significantly is for $\mathbf{MCE}3_2$ where x_{it} is correlated with c_i. As we saw in section 5.3 this was only process for which the estimated APEs of the Heckman method was far off the true values. From table 7 we see that the large bias for the APEs is driven by large biases in estimating ρ. In general the MAEs and RMSEs of the coefficient estimates are comparable across the Heckman method and $\mathbf{W}_4$. When $t = 1$ the Heckman method has slightly lower MAEs and RMSEs.

Comparing the results for coefficient estimates for the different Wooldridge specifications against each other reveal the same pattern as for the estimated APEs. There are few significant differences between the different specifications. In fact the differences are now smaller than for the estimated APEs. Rabe-Hesketh and Skrondal (2013) finds that $\mathbf{W}_1$ through $\mathbf{W}_4$ have very similar performance when estimating ρ a finding I am able to replicate. Furthermore I show that the same holds for β_1.

Table 7: Finite sample coefficient results for Heckman and $\mathbf{W}_4$.

	Heckman						Wooldridge 4 ($\mathbf{W}_4$)					
	ρ			β_1			ρ			β_1		
	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE
	MCE$_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$										
3	-4.79	0.23	0.30	5.74	0.15	0.21	-1.72	0.30	0.38	5.52	0.15	0.21
4	0.88	0.16	0.20	2.38	0.10	0.13	1.81	0.18	0.22	2.45	0.10	0.14
5	-0.21	0.13	0.16	0.89	0.08	0.10	0.11	0.14	0.18	1.07	0.08	0.10
8	0.36	0.08	0.10	0.58	0.05	0.06	0.74	0.09	0.11	0.74	0.05	0.07
12	-0.87	0.07	0.08	0.40	0.04	0.05	-1.09	0.07	0.08	0.52	0.04	0.05
20	1.89	0.05	0.06	-0.37	0.03	0.04	0.89	0.05	0.06	-0.06	0.03	0.04
	MCE$_{1.2}$	$x_{it} \sim \frac{\chi_1^2 - 1}{\sqrt{2}}$										
3	-3.46	0.24	0.31	4.74	0.17	0.23	-1.37	0.30	0.38	4.38	0.18	0.25
4	-0.41	0.17	0.21	2.30	0.12	0.15	2.03	0.18	0.23	2.78	0.12	0.16
5	-0.21	0.13	0.17	0.82	0.09	0.11	0.25	0.14	0.17	0.88	0.09	0.12
8	0.27	0.08	0.11	0.65	0.07	0.08	0.47	0.09	0.11	0.66	0.07	0.08
12	0.64	0.07	0.08	-0.09	0.05	0.06	0.38	0.07	0.08	-0.04	0.05	0.06
20	2.08	0.05	0.06	0.17	0.04	0.05	0.68	0.05	0.06	0.54	0.04	0.05
	MCE$_{1.3}$	$x_{it} = 0.5x_{i,t-1} + \frac{t+24}{10} + \mathrm{U}(\frac{-1}{2}, \frac{1}{2})$										
3	-11.89	0.33	0.41	5.65	0.29	0.38	-5.11	0.35	0.43	1.49	0.38	0.48
4	3.95	0.21	0.26	-1.39	0.20	0.25	4.75	0.22	0.27	-0.53	0.23	0.29
5	1.67	0.17	0.21	1.39	0.17	0.21	1.86	0.17	0.21	1.45	0.18	0.23
8	-1.05	0.11	0.14	1.25	0.10	0.13	-1.47	0.11	0.14	1.51	0.10	0.13
12	1.39	0.10	0.12	-0.12	0.08	0.10	-0.23	0.10	0.12	0.70	0.08	0.10
20	0.87	0.09	0.11	0.16	0.06	0.08	-0.27	0.09	0.11	0.79	0.06	0.08
	MCE$_{2.1}$	$u_{it} \sim \mathrm{Laplace}\left(0, \sqrt{\frac{1}{2}}\right)$										
3	53.90	0.35	0.44	60.80	0.61	0.76	55.86	0.42	0.53	61.51	0.62	0.78
4	53.71	0.30	0.35	51.49	0.52	0.56	54.02	0.31	0.37	52.12	0.52	0.57
5	54.01	0.28	0.33	47.60	0.48	0.50	52.71	0.28	0.33	48.51	0.49	0.52
8	50.90	0.26	0.29	46.07	0.46	0.47	50.41	0.25	0.29	46.75	0.47	0.48
12	51.62	0.26	0.28	45.29	0.45	0.46	49.82	0.25	0.27	46.29	0.46	0.47
20	51.83	0.26	0.27	42.83	0.43	0.43	48.15	0.24	0.25	44.45	0.44	0.45
	MCE$_{2.2}$	$u_{it} \sim \mathrm{U}\left(\frac{-\sqrt{12}}{2}, \frac{\sqrt{12}}{2}\right)$										
3	-12.50	0.22	0.28	-0.72	0.15	0.19	-11.33	0.28	0.37	-0.50	0.16	0.21
4	-3.17	0.15	0.19	-3.07	0.09	0.11	-1.35	0.17	0.21	-3.02	0.09	0.12
5	-7.39	0.14	0.18	-2.80	0.08	0.10	-6.54	0.15	0.18	-2.43	0.08	0.10
8	-7.02	0.09	0.12	-4.12	0.06	0.07	-6.81	0.09	0.12	-4.08	0.06	0.07
12	-6.00	0.07	0.09	-3.59	0.05	0.06	-6.07	0.07	0.09	-3.52	0.05	0.06
20	-4.47	0.05	0.06	-3.77	0.04	0.05	-5.34	0.05	0.06	-3.51	0.04	0.05
	MCE$_{2.3}$	$u_{it} \sim \mathrm{Log}(0,1)$										
3	-46.73	0.28	0.35	-38.89	0.39	0.40	-46.28	0.29	0.38	-39.51	0.40	0.41
4	-41.66	0.23	0.27	-40.16	0.40	0.41	-40.39	0.23	0.28	-39.91	0.40	0.41
5	-39.47	0.20	0.24	-41.05	0.41	0.42	-39.84	0.21	0.24	-41.06	0.41	0.42
8	-40.92	0.21	0.22	-40.68	0.41	0.41	-41.12	0.21	0.22	-40.68	0.41	0.41
12	-40.39	0.20	0.21	-40.67	0.41	0.41	-40.26	0.20	0.21	-40.56	0.41	0.41
20	-41.28	0.21	0.21	-40.86	0.41	0.41	-41.27	0.21	0.21	-40.82	0.41	0.41
	MCE$_{3.1}$	$c_i \sim \Gamma(4, 0.5) - 2$										
3	-2.91	0.21	0.27	4.25	0.14	0.19	0.27	0.28	0.37	4.28	0.16	0.20
4	-0.15	0.16	0.20	1.24	0.10	0.13	1.63	0.18	0.22	1.58	0.11	0.14
5	-0.30	0.12	0.16	0.29	0.08	0.09	0.56	0.14	0.17	0.32	0.08	0.10
8	0.10	0.08	0.11	0.73	0.05	0.06	0.59	0.09	0.11	0.85	0.05	0.06
12	-0.07	0.07	0.08	-0.39	0.04	0.05	-0.03	0.07	0.08	-0.30	0.04	0.05
20	1.12	0.04	0.06	-0.33	0.03	0.04	0.36	0.04	0.06	-0.13	0.03	0.04
	MCE$_{3.2}$	$x_{it} \sim \mathcal{N}\left(\frac{c_i}{3}, 1\right)$										
3	97.21	0.49	0.54	5.45	0.11	0.15	-1.35	0.29	0.37	5.41	0.16	0.22
4	80.74	0.41	0.46	7.11	0.10	0.13	0.75	0.19	0.24	2.38	0.11	0.13
5	66.32	0.34	0.38	8.49	0.10	0.13	-0.10	0.14	0.17	1.77	0.09	0.11
8	38.74	0.20	0.24	8.08	0.09	0.11	-0.53	0.10	0.12	0.48	0.06	0.07
12	24.62	0.13	0.16	6.42	0.07	0.08	-0.19	0.07	0.09	0.18	0.04	0.05
20	14.55	0.08	0.10	4.45	0.05	0.06	-0.43	0.05	0.07	0.32	0.03	0.04
	MCE$_{3.3}$	$x_{1,it} \sim \mathcal{N}(0,1)$, $x_{2,it} \sim \mathcal{N}(0,1)$, $\mathrm{Corr}(x_{1,it}, x_{2,it}) = 0.3$										
3	-31.67	0.24	0.31	-1.60	0.12	0.16	-31.95	0.28	0.37	-1.68	0.13	0.17
4	-27.91	0.18	0.23	-4.76	0.09	0.12	-27.81	0.20	0.24	-4.46	0.09	0.12
5	-26.40	0.16	0.20	-6.29	0.09	0.10	-25.47	0.16	0.20	-6.07	0.09	0.11
8	-27.58	0.14	0.17	-5.45	0.07	0.08	-27.76	0.15	0.17	-5.41	0.07	0.08
12	-27.47	0.14	0.16	-5.85	0.06	0.07	-27.32	0.14	0.16	-5.79	0.06	0.07
20	-28.01	0.14	0.15	-5.87	0.06	0.07	-28.03	0.14	0.15	-5.82	0.06	0.07

Table 8: Finite sample coefficient results for $\mathbf{W}_1$ and $\mathbf{W}_2$.

	Wooldridge 1 ($\mathbf{W}_1$)						Wooldridge 2 ($\mathbf{W}_2$)					
	ρ			β_1			ρ			β_1		
	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE
	MCE$_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$										
3	-1.72	0.30	0.38	5.52	0.15	0.21	-1.72	0.30	0.37	5.56	0.15	0.20
4	1.81	0.18	0.22	2.45	0.10	0.14	2.07	0.18	0.22	2.38	0.10	0.14
5	0.11	0.14	0.18	1.07	0.08	0.10	0.18	0.14	0.18	1.02	0.08	0.10
8	0.74	0.09	0.11	0.74	0.05	0.07	0.80	0.09	0.11	0.72	0.05	0.07
12	-1.09	0.07	0.08	0.52	0.04	0.05	-1.09	0.07	0.08	0.49	0.04	0.05
20	0.89	0.05	0.06	-0.06	0.03	0.04	0.90	0.05	0.06	-0.08	0.03	0.04
	MCE$_{1.2}$	$x_{it} \sim \frac{\chi_1^2 - 1}{\sqrt{2}}$										
3	-1.37	0.30	0.38	4.38	0.18	0.25	-1.89	0.30	0.38	4.23	0.18	0.25
4	2.03	0.18	0.23	2.78	0.12	0.16	2.07	0.18	0.23	2.65	0.12	0.16
5	0.25	0.14	0.17	0.88	0.09	0.12	0.41	0.14	0.17	0.76	0.09	0.12
8	0.47	0.09	0.11	0.66	0.07	0.08	0.53	0.09	0.11	0.58	0.07	0.08
12	0.38	0.07	0.08	-0.04	0.05	0.06	0.37	0.07	0.08	-0.11	0.05	0.06
20	0.68	0.05	0.06	0.54	0.04	0.05	0.66	0.05	0.06	0.51	0.04	0.05
	MCE$_{1.3}$	$x_{it} = 0.5x_{i,t-1} + \frac{t+24}{10} + \mathrm{U}(\frac{-1}{2}, \frac{1}{2})$										
3	-5.11	0.35	0.43	1.49	0.38	0.48	-5.53	0.35	0.43	1.50	0.38	0.48
4	4.75	0.22	0.27	-0.53	0.23	0.29	4.84	0.22	0.27	-0.55	0.23	0.29
5	1.86	0.17	0.21	1.45	0.18	0.23	1.95	0.17	0.21	1.37	0.18	0.23
8	-1.47	0.11	0.14	1.51	0.10	0.13	-1.44	0.11	0.14	1.50	0.10	0.13
12	-0.23	0.10	0.12	0.70	0.08	0.10	-0.22	0.10	0.12	0.69	0.08	0.10
20	-0.27	0.09	0.11	0.79	0.06	0.08	-0.26	0.09	0.11	0.77	0.06	0.08
	MCE$_{2.1}$	$u_{it} \sim \mathrm{Laplace}\left(0, \sqrt{\frac{1}{2}}\right)$										
3	55.86	0.42	0.53	61.51	0.62	0.78	55.37	0.41	0.52	60.71	0.61	0.77
4	54.02	0.31	0.37	52.12	0.52	0.57	53.81	0.31	0.37	51.78	0.52	0.57
5	52.71	0.28	0.33	48.51	0.49	0.52	52.42	0.28	0.33	48.18	0.48	0.51
8	50.41	0.25	0.29	46.75	0.47	0.48	50.45	0.25	0.29	46.55	0.47	0.48
12	49.82	0.25	0.27	46.29	0.46	0.47	49.74	0.25	0.27	46.15	0.46	0.47
20	48.15	0.24	0.25	44.45	0.44	0.45	48.10	0.24	0.25	44.36	0.44	0.45
	MCE$_{2.2}$	$u_{it} \sim \mathrm{U}\left(\frac{-\sqrt{12}}{2}, \frac{\sqrt{12}}{2}\right)$										
3	-11.33	0.28	0.37	-0.50	0.16	0.21	-11.52	0.28	0.36	-0.60	0.16	0.20
4	-1.35	0.17	0.21	-3.02	0.09	0.12	-1.26	0.17	0.21	-3.02	0.09	0.12
5	-6.54	0.15	0.18	-2.43	0.08	0.10	-6.43	0.15	0.18	-2.47	0.08	0.10
8	-6.81	0.09	0.12	-4.08	0.06	0.07	-6.71	0.09	0.12	-4.08	0.06	0.07
12	-6.07	0.07	0.09	-3.52	0.05	0.06	-6.02	0.07	0.09	-3.52	0.05	0.06
20	-5.34	0.05	0.06	-3.51	0.04	0.05	-5.33	0.05	0.06	-3.52	0.04	0.05
	MCE$_{2.3}$	$u_{it} \sim \mathrm{Log}(0,1)$										
3	-46.28	0.29	0.38	-39.51	0.40	0.41	-46.79	0.30	0.38	-39.47	0.40	0.41
4	-40.39	0.23	0.28	-39.91	0.40	0.41	-40.43	0.22	0.28	-39.91	0.40	0.41
5	-39.84	0.21	0.24	-41.06	0.41	0.42	-39.81	0.21	0.24	-41.07	0.41	0.42
8	-41.12	0.21	0.22	-40.68	0.41	0.41	-41.12	0.21	0.22	-40.69	0.41	0.41
12	-40.26	0.20	0.21	-40.56	0.41	0.41	-40.25	0.20	0.21	-40.56	0.41	0.41
20	-41.27	0.21	0.21	-40.82	0.41	0.41	-41.27	0.21	0.21	-40.83	0.41	0.41
	MCE$_{3.1}$	$c_i \sim \Gamma(4, 0.5) - 2$										
3	0.27	0.28	0.37	4.28	0.16	0.20	0.14	0.28	0.36	4.34	0.15	0.20
4	1.63	0.18	0.22	1.58	0.11	0.14	1.22	0.18	0.22	1.55	0.11	0.14
5	0.56	0.14	0.17	0.32	0.08	0.10	0.71	0.14	0.17	0.31	0.08	0.10
8	0.59	0.09	0.11	0.85	0.05	0.06	0.71	0.09	0.11	0.83	0.05	0.06
12	-0.03	0.07	0.08	-0.30	0.04	0.05	-0.01	0.07	0.08	-0.31	0.04	0.05
20	0.36	0.04	0.06	-0.13	0.03	0.04	0.37	0.04	0.06	-0.13	0.03	0.04
	MCE$_{3.2}$	$x_{it} \sim \mathcal{N}\left(\frac{c_i}{3}, 1\right)$										
3	-1.35	0.29	0.37	5.41	0.16	0.22	-1.35	0.29	0.37	5.43	0.16	0.22
4	0.75	0.19	0.24	2.38	0.11	0.13	0.70	0.19	0.24	2.35	0.11	0.13
5	-0.10	0.14	0.17	1.77	0.09	0.11	-0.13	0.14	0.17	1.77	0.09	0.11
8	-0.53	0.10	0.12	0.48	0.06	0.07	-0.51	0.10	0.12	0.48	0.06	0.07
12	-0.19	0.07	0.09	0.18	0.04	0.05	-0.17	0.07	0.09	0.18	0.04	0.05
20	-0.43	0.05	0.07	0.32	0.03	0.04	-0.43	0.05	0.07	0.32	0.03	0.04
	MCE$_{3.3}$	$x_{1,it} \sim \mathcal{N}(0,1)$, $x_{2,it} \sim \mathcal{N}(0,1)$, $\mathrm{Corr}(x_{1,it}, x_{2,it}) = 0.3$										
3	-31.95	0.28	0.37	-1.68	0.13	0.17	-32.63	0.29	0.37	-1.57	0.13	0.17
4	-27.81	0.20	0.24	-4.46	0.09	0.12	-27.84	0.20	0.24	-4.42	0.09	0.12
5	-25.47	0.16	0.20	-6.07	0.09	0.11	-25.51	0.16	0.20	-6.07	0.09	0.11
8	-27.76	0.15	0.17	-5.41	0.07	0.08	-27.74	0.15	0.17	-5.40	0.07	0.08
12	-27.32	0.14	0.16	-5.79	0.06	0.07	-27.31	0.14	0.16	-5.78	0.06	0.07
20	-28.03	0.14	0.15	-5.82	0.06	0.07	-28.02	0.14	0.15	-5.82	0.06	0.07

Table 9: Finite sample coefficient results for $\mathbf{W}_3$ and $\mathbf{W}_5$.

	Wooldridge 3 ($\mathbf{W}_3$)						Wooldridge 5 ($\mathbf{W}_5$)					
	ρ			β_1			ρ			β_1		
	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE	Average	MAE	RMSE
$\mathbf{MCE}_{1.1}$	$x_{it} \sim \mathcal{N}(0,1)$											
3	-2.58	0.28	0.36	5.12	0.15	0.20	-2.18	0.28	0.35	5.15	0.15	0.20
4	1.63	0.18	0.22	2.14	0.10	0.13	1.99	0.18	0.22	2.07	0.10	0.13
5	-0.01	0.14	0.17	0.84	0.08	0.10	0.15	0.14	0.17	0.79	0.08	0.10
8	0.57	0.08	0.11	0.61	0.05	0.07	0.61	0.08	0.11	0.58	0.05	0.07
12	-1.20	0.07	0.08	0.42	0.04	0.05	-1.20	0.07	0.08	0.39	0.04	0.05
20	0.80	0.05	0.06	-0.13	0.03	0.04	0.81	0.05	0.06	-0.14	0.03	0.04
$\mathbf{MCE}_{1.2}$	$x_{it} \sim \frac{\chi^2_1 - 1}{\sqrt{2}}$											
3	-1.65	0.29	0.37	3.93	0.18	0.24	-0.38	0.30	0.37	4.43	0.18	0.24
4	1.63	0.18	0.22	2.44	0.12	0.16	2.00	0.18	0.23	2.42	0.12	0.16
5	0.12	0.14	0.17	0.61	0.09	0.12	0.37	0.14	0.17	0.55	0.09	0.12
8	0.12	0.09	0.11	0.48	0.07	0.08	0.22	0.09	0.11	0.43	0.07	0.08
12	0.21	0.07	0.08	-0.16	0.05	0.06	0.20	0.07	0.08	-0.22	0.05	0.06
20	0.55	0.05	0.06	0.45	0.04	0.05	0.54	0.05	0.06	0.43	0.04	0.05
$\mathbf{MCE}_{1.3}$	$x_{it} = 0.5x_{i,t-1} + \frac{t+24}{10} + \mathrm{U}(\frac{-1}{2}, \frac{1}{2})$											
3	-6.96	0.35	0.43	1.38	0.38	0.48	-4.51	0.34	0.42	1.48	0.38	0.48
4	3.59	0.21	0.27	-1.18	0.23	0.29	4.89	0.21	0.27	-0.66	0.23	0.29
5	1.11	0.17	0.21	0.68	0.18	0.22	2.00	0.17	0.21	1.30	0.18	0.23
8	-1.66	0.11	0.14	0.95	0.10	0.13	-1.42	0.11	0.14	1.57	0.10	0.13
12	-0.39	0.10	0.12	0.20	0.08	0.10	-0.30	0.10	0.12	0.67	0.08	0.10
20	-0.47	0.09	0.11	0.46	0.06	0.08	-0.41	0.09	0.11	0.80	0.06	0.08
$\mathbf{MCE}_{2.1}$	$u_{it} \sim \mathrm{Laplace}\left(0, \sqrt{\frac{1}{2}}\right)$											
3	55.25	0.41	0.51	61.06	0.61	0.78	55.27	0.40	0.49	60.01	0.60	0.76
4	53.58	0.30	0.37	51.69	0.52	0.56	53.70	0.31	0.37	51.25	0.51	0.56
5	52.48	0.28	0.33	48.18	0.48	0.51	52.29	0.28	0.33	47.84	0.48	0.51
8	50.07	0.25	0.28	46.56	0.47	0.48	50.18	0.25	0.28	46.36	0.46	0.48
12	49.63	0.25	0.27	46.17	0.46	0.47	49.58	0.25	0.27	46.03	0.46	0.47
20	48.03	0.24	0.25	44.38	0.44	0.45	48.00	0.24	0.25	44.29	0.44	0.45
$\mathbf{MCE}_{2.2}$	$u_{it} \sim \mathrm{U}\left(\frac{-\sqrt{12}}{2}, \frac{\sqrt{12}}{2}\right)$											
3	-10.11	0.27	0.34	-1.02	0.16	0.20	-10.51	0.27	0.34	-1.11	0.16	0.20
4	-2.17	0.17	0.21	-3.30	0.10	0.12	-2.08	0.16	0.21	-3.33	0.10	0.12
5	-6.63	0.14	0.18	-2.67	0.08	0.10	-6.45	0.14	0.18	-2.71	0.08	0.10
8	-6.99	0.09	0.12	-4.22	0.06	0.07	-6.91	0.09	0.12	-4.22	0.06	0.07
12	-6.18	0.07	0.09	-3.62	0.05	0.06	-6.13	0.07	0.09	-3.61	0.05	0.06
20	-5.41	0.05	0.06	-3.57	0.04	0.05	-5.40	0.05	0.06	-3.57	0.04	0.05
$\mathbf{MCE}_{2.3}$	$u_{it} \sim \mathrm{Log}(0,1)$											
3	-45.75	0.29	0.37	-39.70	0.40	0.42	-46.29	0.29	0.37	-39.65	0.40	0.42
4	-40.62	0.22	0.28	-40.09	0.40	0.41	-40.75	0.22	0.28	-40.08	0.40	0.41
5	-39.93	0.21	0.24	-41.19	0.41	0.42	-39.89	0.21	0.24	-41.19	0.41	0.42
8	-41.17	0.21	0.23	-40.77	0.41	0.41	-41.19	0.21	0.23	-40.78	0.41	0.41
12	-40.33	0.20	0.21	-40.62	0.41	0.41	-40.31	0.20	0.21	-40.62	0.41	0.41
20	-41.32	0.21	0.21	-40.86	0.41	0.41	-41.32	0.21	0.21	-40.86	0.41	0.41
$\mathbf{MCE}_{3.1}$	$c_i \sim \Gamma(4, 0.5) - 2$											
3	-0.44	0.26	0.34	3.95	0.15	0.20	-0.43	0.26	0.34	4.02	0.15	0.20
4	1.14	0.18	0.22	1.33	0.11	0.14	0.67	0.18	0.22	1.32	0.11	0.14
5	0.07	0.14	0.17	0.10	0.08	0.10	0.20	0.13	0.17	0.10	0.08	0.10
8	0.35	0.08	0.11	0.71	0.05	0.06	0.46	0.09	0.11	0.70	0.05	0.06
12	-0.17	0.07	0.08	-0.39	0.04	0.05	-0.15	0.07	0.08	-0.40	0.04	0.05
20	0.30	0.04	0.06	-0.19	0.03	0.04	0.32	0.04	0.06	-0.19	0.03	0.04
$\mathbf{MCE}_{3.2}$	$x_{it} \sim \mathcal{N}\left(\frac{c_i}{3}, 1\right)$											
3	-0.57	0.29	0.36	5.07	0.16	0.22	-0.68	0.29	0.36	5.09	0.16	0.22
4	0.60	0.19	0.24	2.07	0.11	0.13	0.65	0.19	0.24	2.06	0.11	0.13
5	-0.35	0.14	0.17	1.56	0.09	0.11	-0.24	0.14	0.17	1.58	0.09	0.11
8	-0.52	0.10	0.12	0.35	0.06	0.07	-0.47	0.10	0.12	0.34	0.06	0.07
12	-0.33	0.07	0.09	0.10	0.04	0.05	-0.30	0.07	0.09	0.11	0.04	0.05
20	-0.48	0.05	0.07	0.28	0.03	0.04	-0.47	0.05	0.07	0.28	0.03	0.04
$\mathbf{MCE}_{3.3}$	$x_{1,it} \sim \mathcal{N}(0,1),\ x_{2,it} \sim \mathcal{N}(0,1),\ \mathrm{Corr}(x_{1,it}, x_{2,it}) = 0.3$											
3	-32.08	0.28	0.36	-2.07	0.13	0.17	-32.40	0.28	0.36	-1.99	0.13	0.17
4	-27.61	0.19	0.24	-4.75	0.09	0.12	-27.50	0.19	0.24	-4.73	0.09	0.12
5	-25.66	0.16	0.20	-6.26	0.09	0.11	-25.71	0.16	0.20	-6.26	0.09	0.11
8	-27.80	0.15	0.17	-5.53	0.07	0.08	-27.75	0.15	0.17	-5.52	0.07	0.08
12	-27.39	0.14	0.16	-5.86	0.06	0.07	-27.37	0.14	0.16	-5.86	0.06	0.07
20	-28.09	0.14	0.15	-5.87	0.06	0.07	-28.08	0.14	0.15	-5.87	0.06	0.07

D Stata code for $\mathbf{MCE}_{1.1}$

Below I have inserted the Stata code used to carry out $\mathbf{MCE}_{1.1}$.

```
quietly {
clear all

noi display c(seed)
set more off
cd \\tsclient\C\Users\eirieb\Dropbox\empdynamics\Mthesis\stata\MCEs

*******************************
****    Name of MCE process ****
****    MCE_2 (x-std.norm)  ****
*******************************
local timeperiods "3 4 5 8 12 20"
timer on 2
    foreach bigt of local timeperiods {
    ****    Change values of parameters here    ****
    sca N = 200
    sca s = -24          // Start period
    sca T = `bigt'           // End period
    sca b0 = 0           // Intercept
    sca b1 = 1           // Slope of first x
    sca rho = 0.5        // State dependence in latent specification
    sca MC = $MCE        // Number of Monte Carlo Repetitions
    local file "MCE1-2" // Used in programming filenames

    ****    Starting simulations    ****
    display "`c(current_time)' `c(current_date)'
    timer on 1
    scalar start_time="`c(current_time)'"

        tempname sim
        postfile `sim' b1_lpm rho_lpm            /// AB-fe
        B_rho_he B_b1_he B_rho_W1 B_b1_W1 B_rho_W2 B_b1_W2 B_rho_W3 B_b1_W3 B_rho_W4 \*
            *\B_b1_W4 B_rho_W5 B_b1_W5  /// % Bias
        SE_rho_he SE_b1_he SE_rho_W1 SE_b1_W1 SE_rho_W2 SE_b1_W2 SE_rho_W3 SE_b1_W3 \*
            *\SE_rho_W4 SE_b1_W4 SE_rho_W5 SE_b1_W5   /// Squared Error
        AE_rho_he AE_b1_he AE_rho_W1 AE_b1_W1 AE_rho_W2 AE_b1_W2 AE_rho_W3 AE_b1_W3 \*
            *\AE_rho_W4 AE_b1_W4 AE_rho_W5 AE_b1_W5   /// Absolute Errors
        AB_APE_b1_lpmAB SE_APE_b1_lpmAB AB_APE_b1_he SE_APE_b1_he AE_APE_b1_he \*
            *\AE_APE_b1_lpmAB AB_APE_b1_W1 SE_APE_b1_W1 AE_APE_b1_W1 AB_APE_b1_W2 \*
            *\SE_APE_b1_W2 AE_APE_b1_W2 AB_APE_b1_W3 SE_APE_b1_W3 AE_APE_b1_W3 \*
            *\AB_APE_b1_W4 SE_APE_b1_W4 AE_APE_b1_W4  AB_APE_b1_W5 SE_APE_b1_W5 \*
            *\AE_APE_b1_W5 ///   AB, SE, AE of the AP's for b1
        AB_APE_rho_lpmAB SE_APE_rho_lpmAB AE_APE_rho_lpmAB AB_APE_rho_he SE_APE_rho_he \*
            *\AE_APE_rho_he AB_APE_rho_W1 SE_APE_rho_W1 AE_APE_rho_W1 AB_APE_rho_W2 \*
            *\SE_APE_rho_W2 AE_APE_rho_W2 AB_APE_rho_W3 SE_APE_rho_W3 AE_APE_rho_W3 \*
            *\AB_APE_rho_W4 SE_APE_rho_W4 AE_APE_rho_W4  AB_APE_rho_W5 SE_APE_rho_W5 \*
            *\AE_APE_rho_W5 ///  AB, SE, AE of the AP's for rho
        b1_he rho_he he_conv         /// Heckman
```

```
b1_W1 rho_W1 w1_conv          /// W_1
b1_W2 rho_W2 w2_conv          /// W_2
b1_W3 rho_W3 w3_conv          /// W_3
b1_W4 rho_W4 w4_conv          /// W_4
b1_W5 rho_W5 w5_conv          /// W_5
y_mean y_sd b1 rho b0 N T s MC num_imb        /// Parameters of interest
using 'file', replace

    forv i=1/`=MC' {
    display as text "START MCE`i'"
    ****    Creating data set with right size    ****  */
    drop _all
    set obs `=N'
    gen int id =[_n]
    expand T+abs(s)+1
    sort id
    by id: gen byte t=s+([_n]-1)
    xtset id t
    sort id t
    ****    Creating ALL necessary variables     ****
    gen byte y = .               // y variable
    gen proj = .             // Projection of b*x+rho*l.y+c - without u!
    gen G = .                // G(xb+rho*l.y+u), i.e. P(y=1|x)=P(u>-xb-c)G(xb+c)
    gen x = .                // first explanatory variable
    gen c = .                // fixed individual effect (c_i) (UNOBSERVED)
    gen u = .                // random error term (u_it) (UNOBSERVED)


    ****    Generating RHS variables     ****
    replace x =rnormal()
    by id: replace c =rnormal() if t==`=s'
    by id: replace c=c[1]       // Fixed over time
    replace u =rnormal()

    ****    Generating LFS variables     ****
    replace y=rbinomial(1,0.5) if t==`=s'   // Inital period

    forv t=`=s+1'/`=T' {
        replace proj=b0+b1*x+rho*l.y+c if t==`t'
        replace G=normal(proj) if t==`t'
        replace G=0.9999999 if G>=0.9999999 & t==`t'     // solves som precision\*
            *\ issues
        replace G=0.0000001 if G<=0.0000001 & t==`t'
        replace y=rbinomial(1,G) if t==`t'
    }

    ****    Trimming and fixing the dataset ****
    sort id t
    drop if t<1
    by id: gen byte Ly=l.y
    gen x2=x if t!=1
    by id: egen x_avg=mean(x2)
    drop x2
```

```
forv t=1/`=T' {
    by id: gen x`t'=x[`t']
}

by id: gen y1xbar=x_avg*y[1]
by id: gen y1=y[1]
gen PE_rho=normal(b0+b1*x+rho+c)-normal(b0+b1*x+c) if t>1
sum PE_rho, meanonly
scalar APE_rho=r(mean)
gen PE_b1=b1*normalden(b0+b1*x+rho*l.y+c)
sum PE_b1, meanonly
scalar APE_b1=r(mean)
count
scalar tot=r(N)
count if proj>1 | proj<0
scalar num_imb=r(N)/tot
drop proj G c u PE_rho PE_b1

****    Start of regressions etc    ****
sum y                      // Saving values for y to compare total \*
    *\variation etc
sca y_mean=r(mean)
sca y_sd=r(sd)

****    OLS regression   ****
xtabond2 y Ly x , gmm(Ly) ivstyle(x) nolevel
mat est_lpm=e(b)           // Saving OLS results (coeffecients)
scalar b1_lpm=est_lpm[1,2]
scalar rho_lpm=est_lpm[1,1]
scalar AB_APE_rho_lpmAB=(rho_lpm-APE_rho)/APE_rho
scalar SE_APE_rho_lpmAB=(rho_lpm-APE_rho)^2
scalar AE_APE_rho_lpmAB=abs(rho_lpm-APE_rho)
scalar AB_APE_b1_lpmAB=(b1_lpm-APE_b1)/APE_b1
scalar SE_APE_b1_lpmAB=(b1_lpm-APE_b1)^2
scalar AE_APE_b1_lpmAB=abs(b1_lpm-APE_b1)

****    Heckman          ****
set maxiter 40
cap redprob y Ly x (x), i(id) t(t) quadrat(12)
_diparm logitrho, ilogit
scalar he_rho=r(est)
scalar list he_rho
scalar sigma_u2=he_rho/(1-he_rho)
scalar he_conv=e(converged)
scalar rho_he=_b[Ly]
scalar b1_he=_b[x]
scalar B_rho_he=(rho_he-rho)/rho
scalar B_b1_he=(b1_he-b1)/b1
scalar SE_rho_he=(rho_he-rho)^2
scalar SE_b1_he =(b1_he-b1)^2
scalar AE_rho_he=abs(rho_he-rho)
scalar AE_b1_he=abs(b1_he-b1)
```

```
scalar factor=(1+sigma_u2^2)^(-0.5)
predict xb, xb
replace xb=xb*factor
gen xb_1=(_b[Ly]+_b[x]*x+_b[_cons])*factor
gen xb_0=(_b[x]*x+_b[_cons])*factor
gen xbtest=(_b[Ly]*Ly+_b[x]*x+_b[_cons])*factor

gen PE_rho_est=normal(xb_1)-normal(xb_0)
sum PE_rho_est, meanonly
scalar AB_APE_rho_he=(r(mean)-APE_rho)/APE_rho
scalar SE_APE_rho_he=(r(mean)-APE_rho)^2
scalar AE_APE_rho_he=abs(r(mean)-APE_rho)

gen PE_b1_est=(_b[x]*factor)*normalden(xb)
sum PE_b1_est, meanonly
scalar AB_APE_b1_he=(r(mean)-APE_b1)/APE_b1
scalar SE_APE_b1_he=(r(mean)-APE_b1)^2
scalar AE_APE_b1_he=abs(r(mean)-APE_b1)
drop PE_rho_est PE_b1_est xb xb_1 xb_0

****    W_1 **** */
xtprobit y Ly x y1 x2-x`=T'
scalar w1_conv=e(converged)
scalar rho_W1=_b[Ly]
scalar b1_W1=_b[x]
scalar B_rho_W1=(rho_W1-rho)/rho
scalar B_b1_W1=(b1_W1-b1)/b1
scalar SE_rho_W1=(rho_W1-rho)^2
scalar SE_b1_W1=(b1_W1-b1)^2
scalar AE_rho_W1=abs(rho_W1-rho)
scalar AE_b1_W1=abs(b1_W1-b1)

scalar factor=sqrt(1+e(sigma_u)^2)^(-0.5)
predict xb, xb
gen xb_1=(xb-_b[Ly]*Ly+_b[Ly])*factor
gen xb_0=(xb-_b[Ly]*Ly)*factor
replace xb=xb*factor
gen PE_rho_est=normal(xb_1)-normal(xb_0)
sum PE_rho_est, meanonly
scalar AB_APE_rho_W1=(r(mean)-APE_rho)/APE_rho
scalar SE_APE_rho_W1=(r(mean)-APE_rho)^2
scalar AE_APE_rho_W1=abs(r(mean)-APE_rho)

gen PE_b1_est=(_b[x]*factor)*normalden(xb)
sum PE_b1_est, meanonly
scalar AB_APE_b1_W1=(r(mean)-APE_b1)/APE_b1
scalar SE_APE_b1_W1=(r(mean)-APE_b1)^2
scalar AE_APE_b1_W1=abs(r(mean)-APE_b1)
drop PE_rho_est PE_b1_est xb xb_1 xb_0

****    W_2 ****
xtprobit y Ly x y1 x1-x`=T'
scalar w2_conv=e(converged)
```

```
scalar rho_W2=_b[Ly]
scalar b1_W2=_b[x]
scalar B_rho_W2=(rho_W2-rho)/rho
scalar B_b1_W2=(b1_W2-b1)/b1
scalar SE_rho_W2=(rho_W2-rho)^2
scalar SE_b1_W2=(b1_W2-b1)^2
scalar AE_rho_W2=abs(rho_W2-rho)
scalar AE_b1_W2=abs(b1_W2-b1)

scalar factor=sqrt(1+e(sigma_u)^2)^(-0.5)
predict xb, xb
gen xb_1=(xb-_b[Ly]*Ly+_b[Ly])*factor
gen xb_0=(xb-_b[Ly]*Ly)*factor
replace xb=xb*factor
gen PE_rho_est=normal(xb_1)-normal(xb_0)
sum PE_rho_est, meanonly
scalar AB_APE_rho_W2=(r(mean)-APE_rho)/APE_rho
scalar SE_APE_rho_W2=(r(mean)-APE_rho)^2
scalar AE_APE_rho_W2=abs(r(mean)-APE_rho)

gen PE_b1_est=(_b[x]*factor)*normalden(xb)
sum PE_b1_est, meanonly
scalar AB_APE_b1_W2=(r(mean)-APE_b1)/APE_b1
scalar SE_APE_b1_W2=(r(mean)-APE_b1)^2
scalar AE_APE_b1_W2=abs(r(mean)-APE_b1)
drop PE_rho_est PE_b1_est xb xb_1 xb_0


****    W_3 ****
xtprobit y Ly x y1 x_avg
scalar w3_conv=e(converged)
scalar b1_W3=_b[x]
scalar rho_W3=_b[Ly]
scalar B_rho_W3=(rho_W3-rho)/rho
scalar B_b1_W3=(b1_W3-b1)/b1
scalar SE_rho_W3=(rho_W3-rho)^2
scalar SE_b1_W3=(b1_W3-b1)^2
scalar AE_rho_W3=abs(rho_W3-rho)
scalar AE_b1_W3=abs(b1_W3-b1)

scalar factor=sqrt(1+e(sigma_u)^2)^(-0.5)
predict xb, xb
gen xb_1=(xb-_b[Ly]*Ly+_b[Ly])*factor
gen xb_0=(xb-_b[Ly]*Ly)*factor
replace xb=xb*factor
gen PE_rho_est=normal(xb_1)-normal(xb_0)
sum PE_rho_est, meanonly
scalar AB_APE_rho_W3=(r(mean)-APE_rho)/APE_rho
scalar SE_APE_rho_W3=(r(mean)-APE_rho)^2
scalar AE_APE_rho_W3=abs(r(mean)-APE_rho)

gen PE_b1_est=(_b[x]*factor)*normalden(xb)
sum PE_b1_est, meanonly
```

```
scalar AB_APE_b1_W3=(r(mean)-APE_b1)/APE_b1
scalar SE_APE_b1_W3=(r(mean)-APE_b1)^2
scalar AE_APE_b1_W3=abs(r(mean)-APE_b1)
drop PE_rho_est PE_b1_est xb xb_1 xb_0

****    W_4 ****
xtprobit y Ly x y1 x_avg x1
scalar w4_conv=e(converged)
scalar b1_W4=_b[x]
scalar rho_W4=_b[Ly]
scalar B_rho_W4=(rho_W4-rho)/rho
scalar B_b1_W4=(b1_W4-b1)/b1
scalar SE_rho_W4=(rho_W4-rho)^2
scalar SE_b1_W4=(b1_W4-b1)^2
scalar AE_rho_W4=abs(rho_W4-rho)
scalar AE_b1_W4=abs(b1_W4-b1)

scalar factor=sqrt(1+e(sigma_u)^2)^(-0.5)
predict xb, xb
gen xb_1=(xb-_b[Ly]*Ly+_b[Ly])*factor
gen xb_0=(xb-_b[Ly]*Ly)*factor
replace xb=xb*factor
gen PE_rho_est=normal(xb_1)-normal(xb_0)
sum PE_rho_est, meanonly
scalar AB_APE_rho_W4=(r(mean)-APE_rho)/APE_rho
scalar SE_APE_rho_W4=(r(mean)-APE_rho)^2
scalar AE_APE_rho_W4=abs(r(mean)-APE_rho)

gen PE_b1_est=(_b[x]*factor)*normalden(xb)
sum PE_b1_est, meanonly
scalar AB_APE_b1_W4=(r(mean)-APE_b1)/APE_b1
scalar SE_APE_b1_W4=(r(mean)-APE_b1)^2
scalar AE_APE_b1_W4=abs(r(mean)-APE_b1)
drop PE_rho_est PE_b1_est xb xb_1 xb_0

****    W_5 ****
xtprobit y Ly x y1 x_avg x1 y1xbar
scalar w5_conv=e(converged)
scalar b1_W5=_b[x]
scalar rho_W5=_b[Ly]
scalar B_rho_W5=(rho_W5-rho)/rho
scalar B_b1_W5=(b1_W5-b1)/b1
scalar SE_rho_W5=(rho_W5-rho)^2
scalar SE_b1_W5=(b1_W5-b1)^2
scalar AE_rho_W5=abs(rho_W5-rho)
scalar AE_b1_W5=abs(b1_W5-b1)

scalar factor=sqrt(1+e(sigma_u)^2)^(-0.5)
predict xb, xb
gen xb_1=(xb-_b[Ly]*Ly+_b[Ly])*factor
gen xb_0=(xb-_b[Ly]*Ly)*factor
replace xb=xb*factor
gen PE_rho_est=normal(xb_1)-normal(xb_0)
```

```
        sum PE_rho_est, meanonly
        scalar AB_APE_rho_W5=(r(mean)-APE_rho)/APE_rho
        scalar SE_APE_rho_W5=(r(mean)-APE_rho)^2
        scalar AE_APE_rho_W5=abs(r(mean)-APE_rho)

        gen PE_b1_est=(_b[x]*factor)*normalden(xb)
        sum PE_b1_est, meanonly
        scalar AB_APE_b1_W5=(r(mean)-APE_b1)/APE_b1
        scalar SE_APE_b1_W5=(r(mean)-APE_b1)^2
        scalar AE_APE_b1_W5=abs(r(mean)-APE_b1)
        drop PE_rho_est PE_b1_est xb xb_1 xb_0

        ****    Saving output       ****
        post `sim' (b1_lpm) (rho_lpm)      /// lpmAB
        (B_rho_he) (B_b1_he) (B_rho_W1) (B_b1_W1) (B_rho_W2) (B_b1_W2) (B_rho_W3) (\*
            *\B_b1_W3) (B_rho_W4) (B_b1_W4) (B_rho_W5) (B_b1_W5) /// % Bias
        (SE_rho_he) (SE_b1_he) (SE_rho_W1) (SE_b1_W1) (SE_rho_W2) (SE_b1_W2) (\*
            *\SE_rho_W3) (SE_b1_W3) (SE_rho_W4) (SE_b1_W4) (SE_rho_W5) (SE_b1_W5) \*
            *\/// Squared Error
        (AE_rho_he) (AE_b1_he) (AE_rho_W1) (AE_b1_W1) (AE_rho_W2) (AE_b1_W2) (\*
            *\AE_rho_W3) (AE_b1_W3) (AE_rho_W4) (AE_b1_W4) (AE_rho_W5) (AE_b1_W5) \*
            *\/// Absolute Errors
        (AB_APE_b1_lpmAB) (SE_APE_b1_lpmAB) (AB_APE_b1_he) (SE_APE_b1_he) (\*
            *\AE_APE_b1_he) (AE_APE_b1_lpmAB) (AB_APE_b1_W1) (SE_APE_b1_W1) (\*
            *\AE_APE_b1_W1) (AB_APE_b1_W2) (SE_APE_b1_W2) (AE_APE_b1_W2) (\*
            *\AB_APE_b1_W3) (SE_APE_b1_W3) (AE_APE_b1_W3) (AB_APE_b1_W4) (\*
            *\SE_APE_b1_W4) (AE_APE_b1_W4) (AB_APE_b1_W5) (SE_APE_b1_W5) (\*
            *\AE_APE_b1_W5) ///   AB, SE, AE of the AP's for b1
        (AB_APE_rho_lpmAB) (SE_APE_rho_lpmAB) (AE_APE_rho_lpmAB) (AB_APE_rho_he) (\*
            *\SE_APE_rho_he) (AE_APE_rho_he) (AB_APE_rho_W1) (SE_APE_rho_W1) (\*
            *\AE_APE_rho_W1) (AB_APE_rho_W2) (SE_APE_rho_W2) (AE_APE_rho_W2) (\*
            *\AB_APE_rho_W3) (SE_APE_rho_W3) (AE_APE_rho_W3) (AB_APE_rho_W4) (\*
            *\SE_APE_rho_W4) (AE_APE_rho_W4) (AB_APE_rho_W5) (SE_APE_rho_W5) (\*
            *\AE_APE_rho_W5) /// AB, SE, AE of the AP's for rho
        (b1_he) (rho_he) (he_conv)  /// Heckman
        (b1_W1) (rho_W1) (w1_conv)  /// W_1
        (b1_W2) (rho_W2) (w2_conv)  /// W_2
        (b1_W3) (rho_W3) (w3_conv)  /// W_3
        (b1_W4) (rho_W4) (w4_conv)  /// W_4
        (b1_W5) (rho_W5) (w5_conv)  /// W_5
        (y_mean) (y_sd) (b1) (rho) (b0) (N) (T) (s) (MC) (num_imb)  // Parameters \*
            *\of interest
        noi display "MCE `i', T=`bigt'"
    }


postclose `sim'
use `file', clear
save "results/`file'T=`=T'", replace // to get RMSE's, calculate the root of the \*
    *\MSE.
order b1_lpm rho_lpm y_mean y_sd b1 rho b0 N T s MC he_conv w*, last
noi sum, sep(4)
noi display as text "Monte Carlo: `c(filename)'"
```

```
    noi display as text "Started at " start_time as text", ended at `c(current_time)'"
    timer off 1
    timer list
    noi display "Which gives a total of "r(t1)/60 " minutes"
     }
}
timer off 2
timer list
noi display "Total run time="r(t2)
global seed=c(seed)
noi display c(seed)
```