Estimation of binary choice models using repeated cross sections
(*preliminary draft*)

Jaya Krishnakumar and Thierno Balde

February 2015

Abstract

In this paper we discuss the estimation of a binary choice model with individual effects using a time series of independent cross-sections. We propose a new approach to parameterize the individual effects accounting for both 'cohort effects' that could be potentially correlated with the regressors, as well as purely individual effects. Drawing on Mundlak's (1978) approach, and extending it to suit our setting, we express the 'cohort effect' as a linear function of the means of the explanatory variables. In a first setting, we assume that individuals' choices are not observed and derive the probability that a certain number of individuals choose '1' among the total number of individuals in a given cohort. Then we go on to the special case in which the individual choices are assumed to be observed. Based on the probabilities of cohort means, we estimate the model using the maximum likelihood method and implement it using a heuristic optimization technique (genetic algorithm). Finally, we carry out Monte Carlo simulations to analyze the finite-sample properties of our estimators, in terms of both bias and mean squared error (MSE).

Keywords: Binary choice, pseudo-panels, repeated cross-sections, cohort effects, maximum likelihood

JEL Classification Codes: C23, C25, C51

1 Introduction

In this paper we analyze a binary choice model with individual effects in the context of clustered data drawn from repeated cross-sectional data. The issue of non-linearity that is often problematic in panel data models with specific effects is exacerbated in this setting by a cohort-level analysis of the dependent variable due to the absence of true panel data, which are typically preferred in econometric studies.

Binary choice models are widely used today in the fields of economics, social sciences, political science, and also medical research. For example, assume we wish to study the participation of married women in the labor market. In this case, the dependent variable assumes one of two values, one (1) if the married woman is on the labor market and zero (0) if she is not. Researchers in labor economics suggest that the decision to participate is partly a function of observable characteristics, either of the individual, such as the education level and family income, or of the economy, such as the unemployment rate, and partly a function of factors that are not observable by the researcher. If these unobservable effects are correlated with the explanatory variables, the model cannot be identified without resorting to external instruments in a simple cross-sectional setting. However, if they are time invariant, the model can be identified using panel data. In another example, consider presidential elections in the United States, where there are two dominant political parties: Republicans and Democrats. The dependent variable is the choice of which of these two parties to vote for. Let us say that it assumes the value of one (1) if the chosen candidate is a Democrat, and zero (0) for a Republican. This issue has been the focus of much work by the economist Ray Fair of Yale University, with the publication "Econometrics and Presidential Elections," and by other political scientists. Variables that are often used in voter choice models include the individual characteristics of voters, the inflation and unemployment rates, etc.

Despite widespread interest in these models, the binary dependent variable is often not available at the individual level. This is might be because information on individual choices are very costly to obtain or because the data provided to researchers is restricted owing to privacy concerns. Only aggregate data on choices at the group level are typically published. The most common types of aggregate data are sums and proportions. In political science the behavior of the elector in an election is frequently cited as an example. Because of voters' privacy rights, their individual choices are masked and we are only given information on the number of votes obtained by each candidate. In medical research, individual-level hospitalization data are usually protected, and only aggregate data (proportions) are accessible. If the explanatory variables for all individuals in a given group are the same, then aggregated binary choice models are easy to estimate (Greene, 2004; Maddala, 1983). Conversely, if explanatory models assume different values from one individual in a group to the next, Miller and Plantinga (1999) recommend using group means of all variables to estimate the model. In this case, we lose some information by using the mean of the variables. Also, the interpretation of the estimated parameters is not at the individual, but rather at the group, level. Consequently, this approach does not allow us to make inferences at the individual level. A situation in which the dependent variable was aggregated while explanatory variables were available at the individual level was illustrated by the Pennsylvania gubernatorial election in 2006. In this election an incumbent Democratic governor faced off against a black Republican candidate. The data was collected by the "The Inter-University Consortium for Political and Social Research (ICPSR)" using questionnaires that were completed by voters on the day of the election. For each voter that participated in the survey several characteristics such as race, sex, age, etc. were observed. Moreover, Pennsylvania is divided into five (5) geographical districts. For each one we have information on the number of survey participants, the number of electors who voted for the Democratic candidate, and individual characteristics of each participant, but not how he or she voted.

Panel data currently offer a wide variety of benefits for analyzing behavior at the micro level, but they are not available for many countries. Instead, there are annual household surveys that are based on a large sample of the population, such as the "the British Family Expenditure Survey" or "Labor Supply Survey." In the case of these repeated cross-sectional surveys we are unable to follow a specific household over time, as would be required for a true panel. Thus, the estimation methods currently used for analyzing panel data are inapplicable. To address this problem, Deaton (1985) suggests using cohorts to estimate linear models. His approach is based on aggregating individuals or households into cohorts and treating the population means of these cohorts as "individuals." In this fashion, this new panel (called a "pseudo panel") allows us to track a representative sample from the same cohort of individuals or households over time.

The "pseudo panel" approach has not only been used in applied microeconomics, such as for studying income and savings (see, for example, Beach and Finnie, 2004; Bourguignon *et al.*, 2004; Baldini and Mazzaferro, 1999) but also in many areas of research in the social sciences, including healthcare, education, employment, etc. (e.g. Garner *et al.*, 2002; Glied, 2002; Lauer, 2003; Anderson and Hussey, 2000; Weir, 2003). To construct pseudo-panel datasets, cohorts need to be defined on the basis of a certain number of shared characteristics. The control variable has to be constant for each individual at all points in time or the individual will cease to belong to the group. Also, the control variable must be observable for all individuals in the sample: year of birth, sex, level of education, geographical region are all good criteria for the formation of cells (Dargay and Vythoulkas, 1999). In a word, the construction of these cells has significant ramifications for the magnitude of the bias and the variance of the estimators in a sample of finite size (Verbeek and Nijman, 1992).

A problem with the cohort approach is that we have to replace the cohort population means with empirical observations from the samples, creating an issue of measurement error in the variables. In the case of linear models, the classic "within" estimator is biased. Since the variance of measurement errors can be estimated from individual data, Deaton (1985) and Collado (1997) suggest corrections that account for measurement error. Fuller (1987) proposes a more general estimator that includes Deaton's as a special case. Verbeek and Nijman (1992) demonstrate that we can ignore measurement error if we have a large number of observations per cohort.

In the case of linear models with individual effects in a true panel, the classical estimation method consists of transforming the model by using the deviation from the mean to eliminate individual effects (see, for example, Hsiao [1986] or Arellano and Bover [1990]). This method is not relevant to our study, since our model is "strongly" nonlinear. For nonlinear models, Mundlak (1978) and Chamberlain (1984) suggest parameterizing individual effects as a linear function of the explanatory variables. Once again, this will not work for us because we do not have observations on the same individuals over time. Collado (1998) demonstrates how to overcome this difficulty in parameterizing these effects by using cohort means. In her approach, we have a model with measurement error on the variables causing the error term to be correlated with the explanatory variables. The covariance between the errors and the explanatory variables is a function of the variance of the measurement error, which can be estimated from individual data. To be able to estimate her model, Collado imposes a further restriction that the joint distribution of the error terms and the cohort means be normal. In this fashion she is able to estimate her binary choice model using pseudo-maximum likelihood and the minimum distance method. In our analysis we circumvent this problem of parameterization by expressing the individual-specific effect as a function of the cohort-specific effect while still following the general thrust of Mundlak's approach. Subsequently, unlike Collado we estimate a model whose dependent binary variable is an aggregate, but whose explanatory variables are at the individual level, thus capitalizing on all the information available for individuals.

Our paper is organized as follows. In Section 2, we specify the model of our study starting from a binary model with individual effects. In Section 3 we construct the likelihood function for the variable of interest (the aggregate dependent variable) and perform the estimation using maximum likelihood. Section 4 presents the optimization, which uses the genetic algorithm. In Section 5 we report the results of our Monte Carlo simulations. Conclusions are presented in Section 6.

2 Model specification

When studying discrete choice with panel data the following linear model is often postulated for the underlying latent variable:

$$y_{it}^* = x_{it}^{'}\beta + \alpha_i + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T \tag{1}$$

where y_{it}^* is an unobserved variable, x_{it} is a $(k \times 1)$ vector of observed explanatory variables, α_i is the unobserved individual effect, and u_{it} the error term. In specification (1) we start from the assumption that we retain the same individuals over time, i.e. individual i in period 1 is the same person as individual i in period 2, and so forth. If we assume that x_{it}, α_i, u_{it} are pairwise independent we have a random effects model. Conversely, if α_i and x_{it} are correlated, we can adopt the parametric approach developed by

Chamberlain or Mundlak. This reparametrization allows us to establish a relationship between the fixed-effects and the random-effects models, especially in a linear context.

According to the Mundlak approach (1978), the individual effects and the explanatory variables are correlated as follows:

$$\alpha_i = \bar{x}_i^{'}\gamma + w_i \tag{2}$$

where $E(w_i|x_{it}, t = 1, \ldots, T) = 0$ and $\bar{x}_i = \frac{1}{T}(x_{i1} + x_{i2} + \ldots + x_{iT})$.

The approach taken by Chamberlain (1980, 1985) is more general. α_i and x_{it} stand in the following linear relationship:

$$\alpha_i = \sum_{t=1}^{T} x_{it}^{'}\gamma_t + w_i \tag{3}$$

with $E(w_i|x_{it}, t = 1, \ldots, T) = 0$. This last parameterization has its drawbacks, including the fact that it is cumbersome to estimate the γ_t in the presence of a large number of observations.

In order not to confuse with true panel data, we adopt the notation of R. Moffit (1993)[1]. Many other authors have adopted this notation in the literature of pseudo-panel data, for example Ainhoa Oguiza Tovar (2012) [2]. We index the i^{th} individual at time t with $i_{(t)}$. This individual will not be the same from one period to the next. For example, the second individual in period 1, $2_{(1)}$, will not be the same person as the second individual in period 2, $2_{(2)}$. The number of observations can differ from one period to the next, and is indicated by $N(t)$. Thus, we formulate our model as follows:

$$y^*_{i_{(t)}t} = x^{'}_{i_{(t)}t}\beta + \alpha_{i_{(t)}} + u_{i_{(t)}t}, \quad i_{(t)} = 1_{(t)}, \ldots, N_{(t)}, \quad t = 1, \ldots, T, \tag{4}$$

with the following assumptions:

1. Individuals are independent.
2. $E(u_{i_{(t)}t}|x_{i_{(t)}t}) = 0$ for all i and t
3. $\alpha_{i_{(t)}}$ is potentially correlated with expanatory variables $x_{i_{(t)}t}$

Although we allow for the possibility that $\alpha_{i_{(t)}}$ is potentially correlated with explanatory variables $x_{i_{(t)}t}$, in our setting, the Mundlak method presented in (2) and the Chamberlain method in (3) are inapplicable since we do not have observations on the same individuals from one period to the next to be able to estimate the coefficients of these equations. We propose a new solution to circumvent this problem at the cohort level.

[1] R. Moffit (1993), Identification and estimation of dynamic models with a time series of repeated cross-sections.

[2] Ainhoa Oguiza Tovar, Inmaculada Gallastegui Zulaica, Vicente Nunez-Anton (2012), Analysis of pseudo-panel data with dependent samples.

We decompose the individual effects $\alpha_{i(t)}$ into two parts: α^*_{ct} is a "cohort-specific effect" representing the mean of the individual effects of the *population* in cohort c at time t, and $\xi_{i(t)}$ is the deviation of $\alpha_{i(t)}$ from that mean:

$$\alpha_{i(t)} = \alpha^*_{ct} + \underbrace{(\alpha_{i(t)} - \alpha^*_{ct})}_{\xi_{i(t)}} = \alpha^*_{ct} + \xi_{i(t)}. \tag{5}$$

The mean of the individual effects of the *sample* in cohort c at time t, $\bar{\alpha}_{ct}$, can be decomposed as follows: the *population* mean in cohort c at time t, α^*_{ct}, and a deviation from this mean attributable to sampling error in the data, denoted $v_{c(t)}$

$$\bar{\alpha}_{ct} = \alpha^*_{ct} + v_{c(t)}. \tag{6}$$

Substituting (6) into (5) yields:

$$\alpha_{i(t)} = \bar{\alpha}_{ct} - v_{c(t)} + \xi_{i(t)}. \tag{7}$$

If we assume that populations change little from one period to the next (a very important hypothesis for the construction of our model), then the population mean α^*_{ct} is invariant in time, i.e. $\alpha^*_{ct} = \alpha^*_c$. If we further assume that the cohort size is sufficiently large, i.e. $n_{ct} \simeq n_c \to \infty$, then the sampling error tends toward zero ($v_{c(t)} \to 0$). The two preceding assumptions together have the effect that $\alpha^*_c \simeq \bar{\alpha}_c$, and thus

$$\alpha_{i_{(t)}} = \bar{\alpha}_c + \xi_{i_{(t)}}. \tag{8}$$

Now consider the Mundlak (1978) approach:

$$\alpha_{i(t)} = \bar{x}_i^{'}\gamma + w_{i(t)}, \tag{9}$$

with $E(w_{i(t)}|x_{i_{(t)}t},\ t = 1, \ldots, T) = 0$, $w_{i(t)} \sim \text{iid}(0, \sigma^2_w)$ and $\bar{x}_i^{'}$ (average of the observations over time) unobserved. A simple algebraic manipulation of (9) yields:

$$\alpha_{i(t)} = x'_{i_{(t)}t}\gamma + (\bar{x}_i - x_{i_{(t)}t})^{'}\gamma + w_{i(t)}. \tag{10}$$

We can now take the mean of the sample observations for each cohort in each period

$$\underbrace{\frac{1}{n_{ct}}\sum_{i=1}^{n_{ct}}\alpha_{i(t)}}_{\bar{\alpha}_{ct}} = \underbrace{\frac{1}{n_{ct}}\sum_{i=1}^{n_{ct}}x'_{i_{(t)}t}}_{\bar{x}'_{ct}}\gamma + \left(\underbrace{\frac{1}{n_{ct}}\sum_{i=1}^{n_{ct}}\bar{x}_i}_{\bar{x}_{c.}} - \underbrace{\frac{1}{n_{ct}}\sum_{i=1}^{n_{ct}}x_{i_{(t)}t}}_{\bar{x}_{ct}}\right)^{'}\gamma + \underbrace{\frac{1}{n_{ct}}\sum_{i=1}^{n_{ct}}w_{i(t)}}_{\bar{w}_{ct}}$$

$$\bar{\alpha}_{ct} = \bar{x}'_{ct}\gamma + (\bar{x}_c - \bar{x}_{ct})'\gamma + \bar{w}_{ct}.$$

Assume that we can estimate the conditional expectation (or the true mean) $E(x_{i_{(t)}t}|i \in c) = \mu_{ct} = \mu_c\ \forall i, t$ from $\bar{x}_{ct}$. When $n_{ct} \simeq n_c \to \infty$ and for similar populations, we have:

$$\bar{x}_{ct} = \frac{1}{n_c}\sum_{i\in c}x_{i_{(t)}t} \to \mu_{ct} = \mu_c$$

$$\frac{1}{n_c}\sum_{i\in c}\bar{x}_i = \frac{1}{n_c}\frac{1}{T}\sum_{i\in c}\sum_{t}x_{i_{(t)}t} \to \mu_c.$$

Thus,

$$\left(\frac{1}{n_c}\sum_{i\in c}\bar{x}_i - \bar{x}_{ct}\right) \to 0.$$

We also have

$$\bar{w}_{ct} = \frac{1}{n_c}\sum_{i\in c} w_{i_{(t)}t} \to 0,$$

$$\bar{\alpha}_{ct} \to \bar{\alpha}_c.$$

Consequently,

$$\bar{\alpha}_c \simeq \bar{x}_c^{'}\gamma.$$

Substituting all these results into (8),

$$\alpha_{i_{(t)}} = \bar{x}_c^{'}\gamma + \xi_{i_{(t)}}.$$

Thus, our latent variable postulated in (4) becomes:

$$y^*_{i_{(t)}t} = x^{'}_{i_{(t)}t}\beta + \bar{x}_c^{'}\gamma + \xi_{i_{(t)}} + u_{i_{(t)}t}, \tag{11}$$

with $\xi_{i_{(t)}}$ and $u_{i_{(t)}t}$ having expectation and covariance zero (since the individuals change from one period to the next) and being independent of $x_{i_{(t)}t}$ and $\bar{x}_c$. This model is thus valid for large n_{ct}.

The variable $y^*_{i_{(t)}t}$ being latent, we observe:

$$y_{i_{(t)}t} = \begin{cases} 1 & \text{if } y^*_{i_{(t)}t} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Thus,

$$p_{i_{(t)}t} \equiv P(y_{i_{(t)}t} = 1) = Pr(x^{'}_{i_{(t)}t}\beta + \alpha_{i_{(t)}} + u_{i_{(t)}t} > 0). \tag{13}$$

Working from the model in (1) we can use standard estimation techniques, such as maximum likelihood (after making some preliminary assumptions on u_{it} and α_i) to estimate the parameters of the model.

If we do not have a "true" panel, as in (4), the estimates obtained will not be consistent. Collado (1998), drawing on the approach in Chamberlain (1984) for individual effects, shows how to obtain consistent results using cohort averages.

In our model, given the absence of a true panel and possibly the unavailability of the dependent variable at the individual level, we work on the level of aggregates (proportions, sums) while retaining explanatory variables at the individual level. Assume that we have a repeated cross-sectional dataset on binary choices for N individuals with $N = N_{(1)} + N_{(2)} + \ldots + N_{(T)}$. Let cohorts of variable sizes (n_{ct}) and homogeneous. Also, let $N_{(t)} = \sum_{c=1}^{C} n_{ct}$. In each cohort c ($c = 1, \ldots, C$), individual $i_{(t)}$ gives the response $y_{i_{(t)}t} = 1$ with probability ($p_{i_{(t)}t}$), or response $y_{i_{(t)}t} = 0$

with probability $(1 - p_{i_{(t)}t})$. $y_{i_{(t)}t}$ is thus a Bernoulli variable conditional on $p_{i_{(t)}t}$, i.e. $f(y_{i_{(t)}t} \backslash p_{i_{(t)}t}) = (p_{i_{(t)}t})^{y_{i_{(t)}t}}(1 - p_{i_{(t)}t})^{1-y_{i_{(t)}t}}$ where $0 < p_{i_{(t)}t} < 1$. For the reasons given above, we are interested in the cohort mean $\bar{y}_{ct} = \frac{1}{n_{ct}} \sum_{i=1}^{n_{ct}} y_{i_{(t)}t}$, where $0 \leq \bar{y}_{ct} \leq 1$, $c = 1, \ldots, C$. We can establish the table of the distribution of $\bar{y}_{ct}$ as follows:

$\bar{y}_{ct}$	$\frac{0}{n_{ct}} = 0$	$\frac{1}{n_{ct}}$	$\frac{2}{n_{ct}}$	$\ldots$	$\frac{k}{n_{ct}}$	$\ldots$	$\frac{n_{ct}}{n_{ct}} = 1$
$\sum_{i=1}^{n_{ct}} y_{i_{(t)}t}$	0	1	2	$\ldots$	k	$\ldots$	n_{ct}
$P\left(\bar{y}_{ct} = \frac{k}{n_{ct}}\right) = P\left(\sum_{i=1}^{n_{ct}} y_{i_{(t)}t} = k\right)$	P_0	P_1	P_2	$\ldots$	P_k	$\ldots$	$P_{n_{ct}}$

Since the observations are independent (by assumption), we have:

$$\begin{aligned} P_0 = P\left(\bar{y}_{ct} = 0\right) &= P\left(\sum_{i=1}^{n_{ct}} y_{i_{(t)}t} = 0\right) = P\left(y_{i_{(t)}t} = 0,\ \forall i \in c\right) \\ &= p^0_{1_{(t)}t} \ldots p^0_{n_{c_{(t)}}t} = \prod_{i=1}^{n_{ct}} p^0_{i_{(t)}t} = \prod_{i=1}^{n_{ct}} (1 - p^1_{i_{(t)}t}) \end{aligned}$$

where $p^1_{i_{(t)}t}$ is the probability that individual i in cohort c choses one (1) at time t, and $p^0_{i_{(t)}t}$ the probability of the complementary event.

$$\begin{aligned} P_1 &= P\left(\bar{y}_{ct} = \frac{1}{n_{ct}}\right) = P\left(\sum_{i=1}^{n_{ct}} y_{i_{(t)}t} = 1\right) \\ &= P\left(y_{i_{(t)}t} = 1, y_{j_{(t)}t} = 0,\ \forall i \neq j,\ i, j \in c\right) \\ &= p^1_{1_{(t)}t} p^0_{2_{(t)}t} \ldots p^0_{n_{c_{(t)}}t} + p^0_{1_{(t)}t} p^1_{2_{(t)}t} p^0_{3_{(t)}t} \ldots p^0_{n_{c_{(t)}}t} + \ldots + p^0_{1_{(t)}t} \ldots p^0_{n_{c-1_{(t)}}t} p^1_{n_{c_{(t)}}t} \\ &= \sum_{i=1}^{n_{ct}} p^1_{i_{(t)}t} \prod_{j=1, i \neq j}^{n_{ct}} p^0_{j_{(t)}t} \\ &= \sum_{i=1}^{n_{ct}} p^1_{i_{(t)}t} \prod_{j=1, i \neq j}^{n_{ct}} (1 - p^1_{j_{(t)}t}). \end{aligned}$$

Example: Assume that there are three individuals per cohort ($n_{ct} = 3$). What is the probability that one of these three individuals in cohort c is hospitalized at time t.

$$\begin{aligned} P\left(\bar{y}_{ct} = \frac{1}{3}\right) &= Pr(\text{1 of the 3 individuals is hospitalized}) \\ &= Pr(\text{1 hospitalized, 2 and 3 not hospitalized}) \\ &\quad + Pr(\text{2 hospitalized, 1 and 3 not hospitalized}) \\ &\quad + Pr(\text{3 hospitalized, 1 and 2 not hospitalized}) \\ &= [\text{if } u_{i_{(t)}t} \sim N] \\ &= \Phi(x'_{1_{(t)}t}\beta + \alpha_1)[1 - \Phi(x'_{2_{(t)}t}\beta + \alpha_2)][1 - \Phi(x'_{3_{(t)}tt}\beta + \alpha_3)] \\ &\quad + \Phi(x'_{2_{(t)}t}\beta + \alpha_2)[1 - \Phi(x'_{1_{(t)}t}\beta + \alpha_1)][1 - \Phi(x'_{3_{(t)}t}\beta + \alpha_3)] \\ &\quad + \Phi(x'_{3_{(t)}t}\beta + \alpha_3)[1 - \Phi(x'_{1_{(t)}t}\beta + \alpha_1)][1 - \Phi(x'_{2_{(t)}t}\beta + \alpha_2)] \end{aligned}$$

$$
\begin{aligned}
P_2 = P\left(\bar{y}_{ct} = \frac{2}{n_{ct}}\right) &= P\left(\sum_{i=1}^{n_{ct}} y_{i_{(t)}t} = 2\right) \\
&= P\left(y_{i_{1(t)}t} = 1,\, y_{i_{2(t)}t} = 1,\, y_{j_{(t)}t} = 0\, \forall i_1 \neq i_2 \neq j,\ i_1, i_2, j \in c\right) \\
= p^1_{1_{(t)}t} p^1_{2_{(t)}t} p^0_{3_{(t)}t} \cdots p^0_{n_{c_{(t)}}t} &+ p^1_{1_{(t)}t} p^0_{2_{(t)}t} p^1_{3_{(t)}t} p^0_{4_{(t)}t} \cdots p^0_{n_{c_{(t)}}t} + \ldots + p^1_{1_{(t)}t} p^0_{2_{(t)}t} \cdots p^0_{n_{c-1_{(t)}}t} p^1_{n_{c_{(t)}}t} \\
&+ p^0_{1_{(t)}t} p^1_{2_{(t)}t} p^1_{3_{(t)}t} p^0_{4_{(t)}t} \cdots p^0_{n_{c_{(t)}}t} + \ldots + p^0_{1_{(t)}t} p^1_{2_{(t)}t} \cdots p^0_{n_{c-1_{(t)}}t} p^1_{n_{c_{(t)}}t} \\
&+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&+ p^0_{1_{(t)}t} p^0_{2_{(t)}t} \cdots p^1_{n_{c-1_{(t)}}t} p^1_{n_{c_{(t)}}t} \\
= \sum_{i_1=1}^{n_{ct}} \sum_{i_2>i_1} p^1_{i_{1(t)}t} p^1_{i_{2(t)}t} \prod_{j=1, i_1 \neq i_2 \neq j}^{n_{ct}} p^0_{j_{(t)}t} &= \sum_{i_1=1}^{n_{ct}} \sum_{i_2>i_1} p^1_{i_{1(t)}t} p^1_{i_{2(t)}t} \prod_{j=1, i_1 \neq i_2 \neq j}^{n_{ct}} (1 - p^1_{j_{(t)}t}).
\end{aligned}
$$

Thus, by deduction, we can write the generic probability of the cohort's average choice as:

$$
\begin{aligned}
P_k = P\left(\bar{y}_{ct} = \frac{k}{n_{ct}}\right) &= P\left(\sum_{i=1}^{n_{ct}} y_{i_{(t)}t} = k\right) \\
&= \sum_{i_1=1}^{n_{ct}} \sum_{i_2>i_1} \cdots \sum_{i_k>i_{k-1}} p^1_{i_{1(t)}t} \cdots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{ct}} p^0_{j_{(t)}t} \\
&= \sum_{i_1=1}^{n_{ct}} \sum_{i_2>i_1} \cdots \sum_{i_k>i_{k-1}} p^1_{i_{1(t)}t} \cdots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{ct}} (1 - p^1_{j_{(t)}t}).
\end{aligned}
$$

When the data on the choices $Y_{i_{1(t)}t}$ are available, the generic probability becomes:

$$
\begin{aligned}
P_k &= P\left(\sum_{i=1}^{n_{ct}} y_{i_{(t)}t} = k\right) \\
&= p^1_{i_{1(t)}t} \cdots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{ct}} p^0_{j_{(t)}t} \\
&= p^1_{i_{1(t)}t} \cdots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{ct}} (1 - p^1_{j_{(t)}t}).
\end{aligned}
$$

Here we have the likelihood of the mean of the observations on cohort c at time t as a function of the parameters β, the individual explanatory variables $x_{i_{(t)}t}$, and the individual effects $\alpha_{i_{(t)}t}$ expressed in terms of the individual probabilities p_{it}. In this formulation, if we let the probability of choosing one (1) be the same for each individual in a given cohort (even though this is not always the case), we end up with a very common probability distribution: the binomial.

For nonlinear models, and specifically in our case, the structural parameters only provide us with information on the relative magnitude of the change in $E(y_{i_{(t)}t}/x_{i_{(t)}t})$ resulting from a variation in a unit of $x_{i_{(t)}t}$, while the marginal effects provide us with the absolute magnitude of the change. This is why, when estimating binary choice models, we are often interested in:

i. The signs and statistical significance of the coefficients.

ii. The marginal effects. For example, in a Probit model,

$$E(y_{i_{(t)}t}/x_{i_{(t)}t}) = \Phi(x'_{i_{(t)}t}\beta + \bar{x}'_c\gamma)$$

and the marginal effects are computed as

$$ME = \frac{\partial E(y_{i_{(t)}t}/x_{i_{(t)}t})}{\partial x_{i_{(t)}t}} = \left(\beta + \frac{1}{n_{ct}}\gamma\right)\phi\left(x'_{i_{(t)}t}\beta + \bar{x}'_c\gamma\right) \tag{14}$$

for a continuous variable $x_{i_{(t)}t}$. We see that, unlike in the case of the linear model, here the marginal effect is the product of two factors: all the effects of the explanatory variables on the latent variable, as well as the derivative of the normal cumulative function evaluated at point $y^*_{i_{(t)}t}$. Furthermore, if we consider

$$E(y_{i_{(t)}t}/x_{i_{(t)}t}, d_{i_{(t)}t}) = \Phi\left(x'_{i_{(t)}t}\beta + \bar{x}'_c\gamma + \delta d_{i_{(t)}t}\right)$$

the marginal effects are

$$ME = \Phi\left(x'_{i_{(t)}t}\beta + \bar{x}'_c\gamma + \delta) - \Phi(x'_{i_{(t)}t}\beta + \bar{x}'_c\gamma\right) \tag{15}$$

for a discrete variable $d_{i_{(t)}t}$.

3 Maximum likelihood estimator

Assume that we have a "pseudo-panel" of dimension $(\sum_{c=1}^{C}\sum_{t=1}^{T} n_{ct})$, where C is the number of cohorts, n_{ct} the size of the cohort, and T the number of periods. The maximum likelihood estimators of β and γ are the vectors $\hat{\beta}$ and $\hat{\gamma}$ that give the highest probability of obtaining $\{\bar{y}_{11}, \ldots, \bar{y}_{1T}, \ldots, \bar{y}_{C1}, \ldots, \bar{y}_{CT}\}$ conditional on the explanatory individual variables. This joint probability is written as:

$$L(\beta, \lambda; \bar{y}_{11}, \ldots, \bar{y}_{1T}, \ldots, \bar{y}_{C1}, \ldots, \bar{y}_{CT}) = P(\bar{y}_{11}, \ldots, \bar{y}_{1T}, \ldots, \bar{y}_{C1}, \ldots, \bar{y}_{CT}; \beta, \gamma).$$

By construction of the pseudo-panel, the observations are independent of each other, and so the likelihood is:

$$L(\beta, \lambda; \bar{y}_{11}, \ldots, \bar{y}_{1T}, \ldots, \bar{y}_{C1}, \ldots, \bar{y}_{CT}) = \prod_{t=1}^{T} \prod_{c=1}^{C} P(\bar{y}_{ct})$$

$$= \prod_{t=1}^{T} P(\bar{y}_{1t}) P(\bar{y}_{2t}) \ldots P(\bar{y}_{Ct})$$

$$= \prod_{t=1}^{T} \underbrace{\left[\sum_{i_1=1}^{n_{1t}} \sum_{i_2>i_1} \cdots \sum_{i_k>i_{k-1}} p^1_{i_{1(t)}t} \ldots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{1t}} (1 - p^1_{j_{(t)}t}) \right]}_{\text{cohort for } c=1 \text{ at time } t}$$

$$\ldots \underbrace{\left[\sum_{i_1=1}^{n_{Ct}} \sum_{i_2>i_1} \cdots \sum_{i_k>i_{k-1}} p^1_{i_{1(t)}t} \ldots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{Ct}} (1 - p^1_{j_{(t)}t}) \right]}_{\text{cohort for } c=C \text{ at time } t}.$$

To simplify the calculations, and because the function log is monotonic, it is advisable to work with the log-likelihood function. Thus:

$$\log L(\beta, \lambda; y_{11}, \ldots, y_{1T}, \ldots, y_{C1}, \ldots, y_{CT})$$

$$= \sum_{t=1}^{T} [\log P(\bar{y}_{1t}) + \log P(\bar{y}_{2t}) + \ldots + \log P(\bar{y}_{Ct})]$$

$$= \sum_{t=1}^{T} \left\{ \log \underbrace{\left[\sum_{i_1=1}^{n_{1t}} \sum_{i_2>i_1} \cdots \sum_{i_k>i_{k-1}} p^1_{i_{1(t)}t} \ldots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{1t}} (1 - p^1_{j_{(t)}t}) \right]}_{\text{cohort for } c=1 \text{ at time t}} \right.$$

$$\left. + \ldots + \log \underbrace{\left[\sum_{i_1=1}^{n_{Ct}} \sum_{i_2>i_1} \cdots \sum_{i_k>i_{k-1}} p^1_{i_1 t} \ldots p^1_{i_{k(t)}t} \prod_{j=1, i_1 \neq \ldots \neq i_k \neq j}^{n_{Ct}} (1 - p^1_{j_{(t)}t}) \right]}_{\text{cohort for } c=C \text{ at time t}} \right\}$$

When the data $y_{i_{(t)}t}$ on the choices are observed, we can write the likelihood function for the model as:

$$
\begin{aligned}
L(\beta,\lambda;\bar{y}_{11},\ldots,\bar{y}_{1T},\ldots,\bar{y}_{C1},\ldots,\bar{y}_{CT}) &= \prod_{t=1}^{T}\prod_{c=1}^{C} P(\bar{y}_{ct}) \\
&= \prod_{t=1}^{T} P(\bar{y}_{1t})P(\bar{y}_{2t})\ldots P(\bar{y}_{Ct}) \\
&= \prod_{t=1}^{T} \underbrace{\left[p^1_{i_{1(t)}t}\ldots p^1_{i_{k(t)}t}\prod_{j=1,i_1\neq\ldots\neq i_k\neq j}^{n_{1t}}(1-p^1_{j_{(t)}t})\right]}_{\text{cohort for } c=1 \text{ at time } t}\ldots \\
&\qquad \underbrace{\left[p^1_{i_{1(t)}t}\ldots p^1_{i_{k(t)}t}\prod_{j=1,i_1\neq\ldots\neq i_k\neq j}^{n_{Ct}}(1-p^1_{j_{(t)}t})\right]}_{\text{cohort for } c=C \text{ at time } t}.
\end{aligned}
$$

and the logarithm of the likelihood function is given by:

$$
\begin{aligned}
&logL(\beta,\lambda;y_{11},\ldots,y_{1T},\ldots,y_{C1},\ldots,y_{CT}) \\
&\quad= \sum_{t=1}^{T}\left[logP(\bar{y}_{1t})+logP(\bar{y}_{2t})+\ldots+logP(\bar{y}_{Ct})\right] \\
&\quad= \sum_{t=1}^{T}\left\{log \underbrace{\left[p^1_{i_{1(t)}t}\ldots p^1_{i_{k(t)}t}\prod_{j=1,i_1\neq\ldots\neq i_k\neq j}^{n_{1t}}\left(1-p^1_{j_{(t)}t}\right)\right]}_{\text{cohort for } c=1 \text{ at time } t}+\ldots\right. \\
&\qquad \left. +\log \underbrace{\left[p^1_{i_{1(t)}t}\ldots p^1_{i_{k(t)}t}\prod_{j=1,i_1\neq\ldots\neq i_k\neq j}^{n_{Ct}}(1-p^1_{j_{(t)}t})\right]}_{\text{cohort for } c=C \text{ at time } t}\right\}
\end{aligned}
$$

where

$$
p^1_{i_{(t)}t}=\Phi\left(\frac{x'_{i_{(t)}t}\beta+\bar{x}_c{}'\gamma}{\sigma^2_\epsilon}\right) \text{ and } \sigma^2_\epsilon=\sigma^2_\xi+\sigma^2_u.
$$

At this stage, the next step should be to derive the first-order conditions (FOC) of the preceding function with respect to the parameters. However, in light of the complexity of the analytical form of the log-likelihood function, we opt for numerical optimization using gradient-based or heuristic algorithms. If the log-likelihood function is globally concave, we can be certain of a unique maximum (McFadden, 1973), but discrete choice models often do not guarantee concavity of the objective function. This problematic behavior of the log-likelihood function is attributable to the fact that the probabilities on which it is constructed do not generally have a closed form. We easily observe this in the specification of our probabilities $P\left(\bar{y}_{ct}=\frac{k}{n_{ct}}\right)$. Thus, we might have several local maxima.

Gradient-based numerical optimization techniques are often used because they converge rapidly. However, these methods require considerable computational resources for finding the derivatives (first- and second-order partial derivatives), and often create issues with convergence because the determinant of the Hessian is equal, or nearly equal, to zero (0). Another problem with these optimization methods is that they tend to get stuck in local extrema, while the objective function is highly variable.

Alternative stochastic or heuristic optimization methods have been proposed, such as the Genetic algorithm. Unlike gradient-based methods, the Genetic algorithm we use in this paper is able to simultaneously search for several local extrema, reducing the risk of getting stuck in a local maximum. The Genetic algorithm uses the values of the objective function, but not derivatives, to find the optimal solution. Unlike gradient-based optimization methods, this method can be very slow in converging. However, because it is better at finding global maxima in maximum likelihood functions, this method is gaining in popularity. Before describing this optimization method, we briefly examine the regularity conditions.

Regularity Conditions

Under general regularity conditions, we can establish the good properties of our estimators namely consistency, asymptotic normality and efficiency. These conditions are as follows [in our case $\theta = (\beta, \gamma)$]:

H1: The random variables $(\bar{Y}_{ct})_{c=1,...C;t=1,...,T}$ are independent and identically distributed with probability density $P(\bar{y}_{ct}) = P(\bar{Y}_{ct} = \bar{y}_{ct}; \theta)$, $\theta \in \Theta \subset \mathbb{R}^p$.

H2: The parameter space Θ is compact.

H3: The unknown true value θ_0 of the parameter vector is identifiable: $\forall \theta_0 \neq \theta_1, P(\bar{Y}_{ct} = \bar{y}_{ct}; \theta_0) \neq P(\bar{Y}_{ct} = \bar{y}_{ct}; \theta_1), \forall \bar{y}_{ct}$.

H4: The log-likelihood function $logL(\theta; \bar{y}_{11}, \ldots, \bar{y}_{1T}, \ldots, \bar{y}_{C1}, \ldots, \bar{y}_{CT})$ is continuous in θ.

H5: $E_{\theta_0}[logP(\bar{Y}_{ct} = \bar{y}_{ct}; \theta)]$ exists!

H6: $\frac{1}{CT} logL(\theta; \bar{y}_{11}, \ldots, \bar{y}_{1T}, \ldots, \bar{y}_{C1}, \ldots, \bar{y}_{CT}) \xrightarrow{a.s.} E_{\theta_0}[logP(\bar{Y}_{ct} = \bar{y}_{ct}; \theta)]$

While the above conditions are sufficient to show the consistency, we need other assumptions to show that the asymptotic distribution is normal.

H7: $\theta_0 \in Int(\Theta)$ i.e the model is correctly specified in the sense that it contains the true distribution generating the observations.

H8: The log-likelihood function $logL(\theta; \bar{y}_{11}, \ldots, \bar{y}_{1T}, \ldots, \bar{y}_{C1}, \ldots, \bar{y}_{CT})$ is twice continuously differentiable in a neighborhood of θ_0.

H9: Sum and differential operators are interchangeable.

H10: The information matrix

$$I(\theta_0) = E_{\theta_0}[-\frac{\partial^2 logP(\bar{Y}_{ct} = \bar{y}_{ct}; \theta_0)}{\partial\theta\partial\theta'}]$$

exists and non-singular.

Thus the maximum likelihood estimator has asymptotically normal distribution

$$\sqrt{CT}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0))$$

4 The genetic algorithm procedure

Genetic algorithms help us solve problems that are not amenable to solution by analytical or, in general, even algorithmic methods. These algorithms, introduced in the United States in the 1970s by John Holland in the book *Adaptation in Natural and Artificial System* (1975) and further developed by Goldberg (1989) in *Genetic Algorithms in Search, Optimization, and Machine Learning* are optimization algorithms based on the mechanisms of evolution and natural selection. This section picks up on these notions, drawing on the work of Yu-Hsin Liu and Hani S. Mahmassani (1999), "Global maximum likelihood estimation procedure for multinomial probit (MNP) model parameters." To use these algorithms for solving an optimization problem, the variables of interest [in our case $\theta = (\beta, \gamma)$] are coded as substrings comprising a set of genes. These substrings are concatenated into strings (or chromosomes) called the solution. The set of these chromosomes (solutions) is called a generation. At each iteration, a new generation (children or offspring) is created from the "best-adapted" of those segments of the strings that are handed down from the previous generation (solutions).

Here is an example to illustrate how it works.

In the first row of Figure 1 we have 12 genes coded as binary values (0 or 1). In the second row, our two variables of interest (or substrings) consist of six genes each. In the third row we have one long string, made up of two substrings, representing a solution.

Figure 1: A genetic representation of two variables of interest with 12 genes

| 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 12 genes |

| 1 0 0 0 1 1 | 1 0 0 1 1 1 | 2 substrings |

| 1 0 0 0 1 1 1 0 0 1 1 1 | 1 string |

The following procedure allows us to estimate the parameters of our model using genetic algorithms: initialization of the population, evaluation, selection, crossover or mutation, and the algorithm's stopping criteria.

Under this optimization technique, several solutions (called generation one, G=1, or the initial population), which are assumed more or less good, are created at random. The values of our variables of interest (or parameters to be estimated) are calculated for each string. Subsequently, the values of these variables are evaluated from the log-likelihood, or "fitness," function:

$$f_{(r)} = -\log L(y, X, \theta_{(r)}),$$

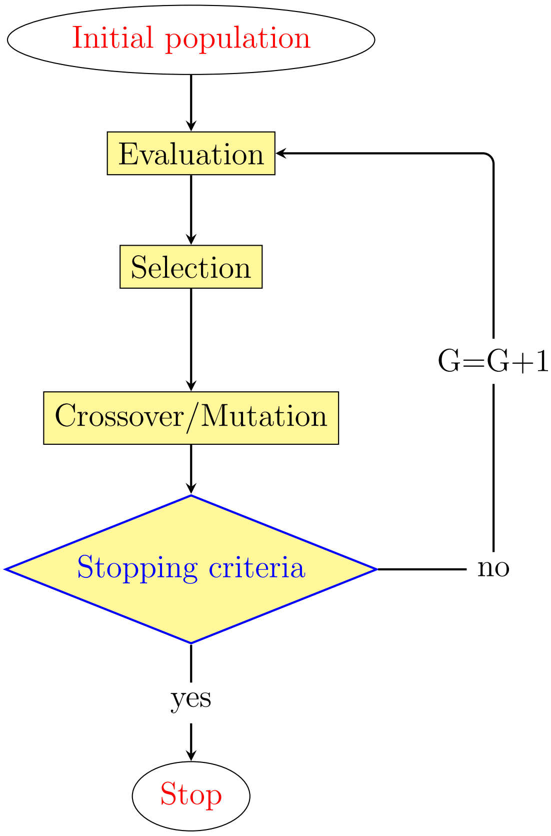
where $f_{(r)}$ is the "fitness" function for string r, $L(y, X, \theta_{(r)})$ is the likelihood function, and $\theta_{(r)}$ is the vector of parameters of the string r. The function $f_{(r)}$ thus allows us to compare all the "strings" of the population. On the basis of this comparison, the strings that give the best "fit," i.e. the most optimal solutions, are selected. These strings, in turn, reproduce and are propagated to the next generation. The population evolves from one generation to the next by crossover[3] of the best strings and causing them to mutate.[4] This iteration continues until the stopping criteria[5] is satisfied, so as to ensure convergence to the optimal solution.

[3]Crossovers yield the best new chromosomes (offspring) by combining the characteristics of the parents. This is, incidentally, one of the most important operations in the process. It occurs with a probability that we have set at 0.8

[4]Mutation consists of a change in the allele (value) of the gene with a low probability (set to $p_m = 0.01$ here). This prevents a premature convergence of the algorithm.

[5]The stopping criteria is often set by the analyst. In our study we set the stopping criteria in terms of capping the number of iterations at $G = 100$.

Figure 2: Structure of a genetic algorithm in an estimation procedure



5 Monte Carlo Study

Monte Carlo simulations are frequently used in pseudo-panel studies to examine the properties of the estimators for finite or infinite populations (asymptotic properties when these distributions are non-standard). In contrast to true panel data, in our context Monte Carlo studies need to account for three (3) dimensions: the number of cohorts (C), the cohort size (n_{ct}), and time (T). In this study we are most interested in asymptotic properties when $n_{ct} \simeq n_c \rightarrow \infty$ because, in practice, we often have C small (owing to the requirement for homogeneity in the construction of the cohorts), T small (because it is often difficult to create large panels), and n_{ct} large.

5.1 Experimental design

We consider the following model (with a single explanatory variable). In order to simplify, we assume that all cohorts are of equal size (n_{ct}=constant)

$$\begin{cases} y^*_{i_{(t)}t} = x_{i_{(t)}t}\beta + \bar{x}'_c\gamma + \xi_{i_{(t)}} + u_{i_{(t)}t} \\ y_{i_{(t)}t} = 1(x_{i_{(t)}t}\beta + \bar{x}'_c\gamma + \xi_{i_{(t)}} + u_{i_{(t)}t} \geq 0) \\ \bar{y}_{ct} = \dfrac{1}{n_{ct}} \sum_{i=1, i\in c}^{n_{ct}} y_{i_{(t)}t}. \end{cases} \tag{16}$$

The data are generated as follows. We start with the explanatory variable $x_{i_{(t)}t}$:

$$x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}. \tag{17}$$

We have three elements on the right-hand side of equation (17). The first element, λ_t, is a parameter that is assumed identical for all individuals at time t but different from one period to the next. This parameter is introduced into the equation to allow for intertemporal variation across cohorts. Using a linear pseudo-panel, Verbeek and Nijman (1992) demonstrate that, in pseudo-panels, large cohorts reduce the bias if λ_t varies over time (t). λ_t is generated from the normal distribution as follows: $\lambda_t \sim N(0, \sigma^2_\lambda)$. The second component, z_i, is a continuous variable identifying the cohort. Assumed independent of λ_t, it is observed for all individuals and invariant over time—otherwise, individuals would be dropping out of their cohort over time. Often, for reasons of tractability, the distribution of the variable z_1 is assumed to be uniform with mean zero (0) and variance one (1), so that the individuals are divided into M intervals, each having the same density function. As we already indicated in the introduction, the most used variable for constructing cohorts in empirical studies is year of birth. We have further assumed that z_i is correlated with the explanatory variable $x_{i_{(t)}t}$, as we see in equation (17). The third element is an error term of the type $E(\eta_{i_{(t)}t}|z_i) = 0$.

To conduct the experiment, we generated observations for $\xi_{i_{(t)}}$, $u_{i_{(t)}t}$ and then for $x_{i_{(t)}t}$ using (17). The generated values of $x_{i_{(t)}t}$ allow us to compute the values of $\bar{x}_c$. Next, we generate the latent variable $y^*_{i_{(t)}t}$, the binary variable $y_{i_{(t)}t}$, and the variable of interest $\bar{y}_{ct}$.

The simulations were envisaged as follows:

(1) $x_{i_{(t)}t}$ is generated as:

$$x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$$

with the following possibilities:

i) $\lambda_t \sim N(0, 1)$ and $\eta_{i_{(t)}t} \sim U(0, 1)$

ii) $\lambda_t \sim N(0, 1)$ and $\eta_{i_{(t)}t} \sim N(0, 1)$

iii) $\lambda_t \sim N(0, 10)$ and $\eta_{i_{(t)}t} \sim U(0, 1)$.

(2) x_c is calculated as follows:

$$x_{ct} = \frac{1}{n_{ct}} \sum_{i=1}^{n_{ct}} x_{i_{(t)}t} \to x_c, \text{when } n_{ct} \simeq n_c \to \infty.$$

(3) Setting $\varepsilon_{i_{(t)}t} = \xi_i + u_{i_{(t)}t}$, we anticipate the following possibilities:

i) $\xi_i \sim N(0, \sigma_\xi^2)$ and $u_{i_{(t)}t} \sim N(0, \sigma_u^2)$ with $\sigma_\varepsilon^2 = \sigma_\xi^2 + \sigma_u^2 = 1$

ii) $\xi_i \sim N(0, \sigma_\xi^2)$ and $u_{i_{(t)}t} \sim$ logistic, with $\sigma_\varepsilon^2 = \sigma_\xi^2 + \sigma_u^2 = 1$.

(4) In consideration of the difficulties obtaining large panels, we set:

i) $T = 3$

ii) $T = 5$.

(5) For the number of cohorts, we consider three cases:

i) $C = 10$

ii) $C = 25$

iii) $C = 50$.

(6) Two principal considerations motivated us to examine all of the following scenarios for n_{ct}. First, we observed that our model is unbiased for n_{ct} large. Second, in empirical studies we often have large cohorts, i.e. $n_{ct} \simeq n_c$ large.

i) $n_{ct} = 10$

ii) $n_{ct} = 25$

iii) $n_{ct} = 50$

iv) $n_{ct} = 100$

v) $n_{ct} = 200$.

(7) To obtain parameter values for the simulations we envisaged three situations:

i) $\beta = 1$, $\gamma = 0.2$ This forms the baseline scenario for our simulations.

ii) $\gamma = 0.2$, we vary β.

iii) $\beta = 1$, we vary γ.
The special case of $\beta = 1$, $\gamma = 0$ allows us to investigate the situation in which there are no individual-specific effects.

We set the number of iterations in each run at 1000.

5.2 Measures of performance

The literature on pseudo-panels provides criteria (statistics) for measuring the performance of an estimator. The first criteria often used is bias. By definition, the bias is the expected difference between the true value of the parameter in question and its estimated value. This measure also reveals the magnitude and the direction of the bias. For example, for a parameter θ:

$$Bias(\hat{\theta}) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta.$$

If all our estimators are unbiased, then the most efficient estimator is the one with the smallest variance. Now, sometimes we have estimators that are weakly biased—in this case the most appropriate metric of their performance is the Mean Square Error (MSE) (Judson and Owen, 1999). By definition, the MSE is the expectation of the squared difference between the true value of the parameter in question and its estimated value. Once again, for a parameter θ, we have:

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = Bias^2(\hat{\theta}) + Var(\hat{\theta}).$$

We see that this criterion simultaneously combines bias and variance, which is why it is so useful.

5.2.1 Structural parameters

If we perform Monte Carlo simulations with $nsim$ iterations, the bias of our estimated parameter becomes $\hat{\beta}$:

$$Bias = \left(\frac{1}{nsim} \sum_{r=1}^{nsim} \hat{\beta}_r \right) - \beta.$$

If we are unable to identify which of the potential estimators has the smallest bias, we use the relative bias (often expressed as a percentage) to compare the performance of the estimators. This bias is computed as follows:

$$RBiais\% = 100 \frac{\left(\frac{1}{nsim} \sum_{r=1}^{nsim} \hat{\beta}_r \right) - \beta}{\beta} = 100 \frac{Bias}{\beta}.$$

The Mean Square Error (MSE) is specified as follows:

$$MSE = \frac{1}{nsim} \sum_{r=1}^{nsim} \left(\hat{\beta}_r - \beta \right)^2.$$

Again, we can compute a relative performance measure expressed in percentages:

$$RMSE\% = 100 \frac{1}{nsim} \sum_{r=1}^{nsim} \left(\frac{\hat{\beta}_r - \beta}{\beta} \right)^2 = 100 \frac{MSE}{\beta^2}.$$

Application of these performance measures to the estimated parameter $\hat{\gamma}$ is similar.

5.2.2 Marginal effects

Marginal effects depend on the individual effects and the values of the explanatory variables at which we wish to evaluate them. In our case, in which the regressors are at the individual level ($x_{i_{(t)}t}$), we have $n_{ct} \times C \times T$ marginal effects for n_{ct} constant . To simplify, we consider the mean of the marginal effects for all observations in the cohort, i.e. the following magnitude:

$$\overline{me}_c = \frac{1}{n_{ct}} \frac{1}{T} \sum_{i=1}^{n_{ct}} \sum_{t=1}^{T} me_{it},$$

where the dimension of $\overline{me}_c$ is $(C \times 1)$, and

$$me_{it} = f(x'_{i_{(t)}t}\beta + \bar{x}'_c\gamma) \left(\beta + \frac{1}{n_{ct}}\gamma \right),$$

where $f()$ is a density function. In our study, $f() \equiv \phi()$, which is the density function of the normal distribution. The values of $\overline{me}_c$ are considered the true values of the marginal effects. In the context of a simulations-based technique, the absolute bias is:

$$Bias(\widehat{\overline{me}}_c) = \left(\frac{1}{nsim} \sum_{r=1}^{nsim} \widehat{\overline{me}}_{c(r)} \right) - \overline{me}_c \tag{18}$$

The relative bias (in percentage) is expressed as:

$$RBias(\widehat{\overline{me}}_c) = 100 \frac{\left(\frac{1}{nsim} \sum_{r=1}^{nsim} \widehat{\overline{me}}_{c(r)} \right) - \overline{me}_c}{\overline{me}_c}. \tag{19}$$

The absolute value of the mean square error (MSE) is:

$$MSE(\widehat{\overline{me}}_c) = \frac{1}{nsim} \sum_{r=1}^{nsim} \left(\widehat{\overline{me}}_{c(r)} - \overline{me}_c \right)^2, \tag{20}$$

and its relative value (in percentage) is:

$$RMSE(\widehat{\overline{me}}_c) = 100 \frac{1}{nsim} \sum_{r=1}^{nsim} \left(\frac{\widehat{\overline{me}}_{c(r)} - \overline{me}_c}{\overline{me}_c} \right)^2, \tag{21}$$

where $\widehat{\overline{me}}_{c(r)}$ is the mean of the estimated marginal effects of all individuals in cohort c at iteration r.

We only retain these two performance measures as we conduct our simulations, even though other measures are potentially available.

5.3 Simulation results

In this section we present the results of our Monte Carlo simulations for various estimators. These results are reported in Tables 1 to 12. The estimators are for the model coefficients, but also for the marginal effects.

1. Tables 1 & 2 present simulation results for the model in the case that variation in the specification is attributable to the size of the cohort n_{ct}.

   As we already discussed in Section 5.2, bias is the name for the difference between the expectation of the estimate and the true parameter value. For example, in Table 1, for $n_{ct} = 10$ the mean bias for $\hat{\beta}$ is -0.2097, indicating that, on average, the deviation of the estimated β from its true value will be approximately 0.2097 in absolute value, with a negative direction. Relative bias, which we also described in Section 5.2, measures the ratio of the bias to the parameter value. This is a good indicator of the magnitude of the bias. Staying with Table 1, for $n_{ct} = 10$, the relative bias of $\hat{\beta}$ is -0.2097, but this time the value indicates that the estimate contains a negative bias amounting to 20.97% of the true parameter value.

   In terms of bias, the estimators $\hat{\beta}$ and $\hat{m}_c$ are underestimated, while $\hat{\gamma}$ is overestimated, regardless of the size of the cohort (n_{ct}). We observe that the relative and absolute biases of the different estimators $\hat{\beta}$, $\hat{\gamma}$ and $\hat{m}_c$ diminish as n_{ct} increases, but these biases persist even when $n_{ct} = 100$. When we move from $n_{ct} = 100$ to $n_{ct} = 200$, the bias approaches zero in the case of our parameter of interest β (because ARB[6] drops from 5.03% to 0.24%) and that of $\hat{\gamma}$ falls precipitously (ARB declines from 40.15% to 4.00%). Thus, we are inclined to believe that the persistence of the bias is attributable to the size of the cohort. For $n_{ct} = 200$, we do not report the results for $\hat{m}_c$ in the tables.[7] We also observe that the biases of the estimates $\hat{m}_c$ are less than those of the estimated coefficients. This might be explained by the presence of the scaling factor $\phi(\cdot)$ in the calculations of the marginal effects. For given explanatory variables X, a strong bias in the coefficients results in small $\phi(\cdot)$, reducing $\beta \cdot \phi(\cdot)$.

   The relative mean squared errors of $\hat{\beta}$, $\hat{\gamma}$, and $\hat{m}_c$ also generally decline when cohort size n_{ct} increases. As of $n_{ct} = 50$ they are already small for $\hat{\beta}$ and $\hat{m}_c$.

   In general, except in the case of $\hat{\gamma}$, we observe that this method performs better for MSE than for bias.

2. Tables 1, 3 and 4 present results of the simulations for the model in which variation in the specification reflects the number of cohorts of C.

   In terms of bias, our results reveal that the absolute value of the relative bias of the three estimators diminishes as the number of cohorts C increases. For example,

[6] ARB stands for Absolute Relative Bias, the absolute value of the relative bias.

[7] This is because the simulations for these values require a great deal of computer memory and processing time.

when $n_{ct} = 10$, the absolute relative bias of $\hat{\beta}$ is $ARB = 20.97\%$ when $C = 10$, $ARB = 17.10\%$ when $C = 25$, and $ARB = 10.91\%$ when $C = 50$. For small n_{ct}, the biases of all three estimators persist even when C increases, while their relative biases are nearly nil when n_{ct} large. It appears that increasing the number of cohorts C leads to underestimating β and m_c and overestimating γ.

As to the MSE, we observe an improvement in the MSEs for $\hat{\beta}$ and $\hat{\gamma}$ when C increases, i.e. their MSE, diminish. For the marginal effects we find the same improvement, except in the cases ($n_{ct} = 50$ and $C : 10 \to 25$).

3. Table 4 presents the results of experiments when the sample size is fixed but the number of cohorts C and the cohort size n_{ct} varies. We know that C and n_{ct} are linked by the formula $N = C \cdot n_{ct}$. Our results thus manifest the trade-off between C and n_{ct} that has been widely discussed in the literature on pseudo-panels and other studies of panels. As we saw above, when T is fixed the performance of the estimators generally depends on the sample size. It is worth noting that these results indicate that greater benefits are conferred by increasing the size (n_{ct}) rather than the number (C), of cohorts. For example, estimating on 10 cohorts with 25 observations per cohort will yield better results than on 25 cohorts with 10 observations per cohort; similarly, 10 cohorts with 50 observations is better than 50 cohorts with 10 observations per cohort.

   This result is of great interest, especially in its application to empirical case studies, because in general the number of cohorts C and the number of periods T are limited and small.

4. Table 5 presents the simulations in which the change is observed in the distribution of the model error term $u_{i_{(t)}t}$.

   Replacing the normal distribution with a logistic distribution for the error term $u_{i_{(t)}t}$ is highly detrimental to the performance of the estimators according to our results. For $\hat{\beta}$, the absolute value of the relative bias rises from 5.03% to 24.16%, an increase of nearly 20%, while its relative MSE increases from 1.40% to 6.93%—a deterioration of 5.53%. In the case of $\hat{\gamma}$, the absolute value of the relative bias increases by nearly 40%, and the MSE by 35%. As to the marginal effects, their relative bias in absolute value deteriorates an average of 16% and their relative MSE by 4%. Conversely, the direction of the bias in both scenarios remains unchanged, i.e. β and m_c remain underestimated and γ overestimated. Nonetheless, there is no reason to expect better results in the second scenario.

5. Table 6 presents the results of simulations following a change in the distribution of the error term $\eta_{i_{(t)}t}$. We know that $\eta_{i_{(t)}t}$ is a component of the data $x_{i_{(t)}t}$. Thus, using a normal distribution $N(0,1)$ instead of a uniform distribution $U(0,1)$ for $\eta_{i_{(t)}t}$ brings greater variability to the data $x_{i_{(t)}t}$. These results reveal that this variability in $x_{i_{(t)}t}$ is beneficial from the perspective of the performance of the estimators, especially for $\hat{\beta}$ and $\hat{m}_c$. For these two estimators, relative absolute

biases and relative MSE become smaller, and even negligible in most cases. For $\hat{\gamma}$, the relative absolute bias and relative MSE also decline, but their magnitude remains non-negligible. With regard to the direction of the bias, it does not change for $\hat{\beta}$, but it does for $\hat{\gamma}$ (becoming negative) and for several $\hat{m}_c$.

6. Table 7 presents the results of the Monte Carlo simulations for the estimators when λ_t is normally distributed but has a greater variance than in the base specification.

   In the context of the estimation of linear pseudo-panels, Verbeek and Nijman (1992) demonstrate the importance of structuring cohorts such that intertemporal variation in the $x_{i_{(t)}t}$ is sufficiently large. According to these authors, satisfying this condition generally allows the bias in the estimators to be reduced. They maintain that this variation usually involves λ_t. The results yielded by our model, which is strongly nonlinear, appear to confirm the results obtained by Verbeek and Nijman in the linear case. Our results are much better when we incorporate more variability in the $x_{i_{(t)}t}$ via λ_t, which is now distributed normally, $N(0, 10)$. For $\hat{\beta}$ (which is the estimate of our principal structural parameter) the absolute value of the relative bias and the relative MSE fall below 1%. In the case of $\hat{\gamma}$ we also find a decline in the absolute value of the relative bias and the relative MSE, but not to the point of being considered small. The relative bias and relative MSEs in the marginal effects also improve. The sign of the bias remains unchanged in these two scenarios. Thus, we here have a very important result, especially in a situation in which it is no easy matter to improve the performance of the estimators—such as nonlinear models.

7. The results reported in Table 8 reflect modifications to the value of β used to generate the data.

   For different values of β our results reveal that the absolute value of the relative bias of $\hat{\beta}$ ranges between 1.40 to 5.03%, that of $\hat{\gamma}$ between 27.35 and 40.15%, and those of the marginal effects between 3 and 6%, on average.

   With regard to the relative MSE: For $\hat{\beta}$ it is between 1.40 and 3.96%, for $\hat{\gamma}$ it falls in the interval 52.75 to 66.75%, and for $\hat{m}_c$ its value averages from 0.20 to 2%.

   On the basis of these results we are unable to detect a direction of movement for the value (increasing or decreasing) of the relative bias or relative MSE with regard to the different values of β, nor can we determine whether any particular scenario dominates the others in terms of performance.

8. Table 9 presents the results of Monte Carlo estimations for data generated using different values of γ. It is of interest here to note that γ represents the degree of correlation between the explanatory variables $x_{i_{(t)}t}$ and the individual effects $\alpha_{i_{(t)}}$.

   As in the previous case, it is impossible here to identify the direction of the magnitude in which the relative bias or the relative MSE moves as a function of the parameter γ. However, unlike in that case, one scenario in particular captures our attention, that of $\beta = 1$ and $\gamma = 0$ (indicating that there is no correlation between $x_{i_{(t)}t}$ and $\alpha_{i_{(t)}}$). This is the situation of "independent random effects." This scenario dovetails nicely with our estimation method in terms of performance. Biases are negligible and the MSE very small for the different estimators. As to the sign of the bias, $\hat{\beta}$ continues to be underestimated, $\hat{\gamma}$ overestimated, and $\hat{m}_c$ underestimated.

9. Table 10 presents Monte Carlo results for the estimators when the pseudo-panel becomes larger.

   For empirical reasons, we would like to have estimation methods that perform well on a sample with small T. When $T = 3$ (which we consider small), we already obtain promising results. Nonetheless, we here consider what happens when T rises to five (5). The absolute values of the relative biases and relative MSEs only change marginally with the change to $T = 5$, but this slight change is an improvement.

9. Table 11 presents the impact of specification error on the estimators. We consider two scenarios.

   First scenario: We generate the model using the values β and γ from the baseline run. However, in the specification of the probabilities (cumulative normal) incorporated into the likelihood, we set $\gamma = 0$, i.e. explanatory variables at the level of the mean are not taken into consideration—meaning that we do not know the correlation between $x_{i_{(t)}t}$ and $\alpha_{i_{(t)}}$. Examination of these results reveals that this specification undermines the performance of the estimators. For example, in the case of $\hat{\beta}$, the absolute value of the relative bias rises from 5.03% to 17.36%, a deterioration of 12.33%. This observed deterioration is not a great surprise, being a result we have already observed in the case of linear models. The relative MSE of $\hat{\beta}$ only deteriorates by approximately 3%. As to the marginal effects, the absolute value of their relative biases increases by an average of 7%, while their relative MSEs deteriorate by an average of approximately 2%.

   Second scenario: The model is generated like in the first scenario. However, in this case, when specifying the individual probabilities we ignore the individual explanatory variables $x_{i_{(t)}t}$, thus retaining only the explanatory variables at the level of the mean, which explicitly translate the correlation between $x_{i_{(t)}t}$ and $\alpha_{i_{(t)}}$. This extreme situation is quite unappealing, because there is no reason to expect our method to perform well—nonetheless, we examine the degree of deterioration. These results confirm our expectations. The relative bias of the estimators snowball, just like the relative MSE. With regard to this latter, for example, for $\hat{\gamma}$ it increases by nearly 162%, while for $\hat{m}_c$ its average deterioration is nearly 157%. Moreover, we observe that while the estimations of the marginal effects are biased downward in the baseline specification, these same estimates are biased upward here. The lesson learned from our results is that it is better to be missing explanatory variables at the level of the mean than at the individual level.

10. Table 12 indicates a loss of efficiency[8] of the estimators when we use grouped data instead of individual data. This loss of efficiency is of the order of 52% about estimator $\hat{\beta}$, of the order of 55% about estimator $\hat{\gamma}$, of the order of 10-22% about estimators of marginal effects $\hat{m}_c$.

By analyzing the simulation results, we believe that the bias of estimators comes from the cohort size because the larger cohort size n_{ct}, the smaller the bias of the estimators.

6 Conclusion

In this paper we have developed a technique for estimating binary choice models with individual effects from aggregate data on choices made by individuals, either to compensate for the absence of individual data or because we wish to analyze the average behavior of cohorts in the absence of a true panel. This is an econometric technique that allows us to recover the parameters that explain individual choices from grouped data on choices. Thanks to these recovered parameters we have also been able to analyze estimations of marginal effects in the framework of binary choice models.

We conducted Monte Carlo simulations to examine the behavior of our estimators with small samples. The results obtained clearly underscore several important points.

For small T and C, we have found that the bias and MSE of the estimators improve as the number of observations per cohort increase, but they do not disappear even when the size reaches 100. At approximately 200 individuals per cohort, the bias vanishes in the estimation of the main structural parameter β, but not for γ. Increasing the number of cohorts C is also key to improving the performance of our estimators. In the case of a sample of a fixed size, it appears to be more helpful to increase the size of the cohort n_{ct} than the number of cohorts C (trade-off in favor of n_{ct}).

Replacing the normal distribution with the logistic distribution for the model error term undermines the performance of the various estimators.

Like in Verbeek and Nijman (1992), our results corroborate that the presence of variability in the explanatory variables improves the estimators, which is perhaps most pronounced in the case of intertemporal variation over the factor λ_t.

Changes to the value of β that generate the data do not appear to significantly affect the bias and MSE of the estimated parameters, except in the case of γ.

Modifications to the value of γ do not have a significant impact on the bias or the MSE of our estimate. Nonetheless, the case in which we assume the absence of correlation

[8] The loss of efficiency of the estimators of grouped data compared individual data is calculated as follows: $\texttt{Efficiency} = \frac{1}{MSE} \rightarrow \texttt{LossEfficiency} = \frac{EfficiencyIndividual}{EfficiencyGrouped}$.

between the explanatory variables and the individual effects dominates the other cases in terms of improvements.

If the cohort size is large, our estimation method performs quite well even for T small ($T = 3$), and this performance improves when the number of periods increases (though this improvement is not across-the-board).

Finally, a poor specification of the probabilities that make up the likelihood exacerbates the bias and the MSE of the estimators. This deterioration snowballs if we do not have explanatory variables at the individual level in the probability equation.

References

Arulampalan, W. and Stewart, M. B. (2007), "Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators", *Discussion Paper No. 3039*, University of Warwick and IZA.

Baltagi, B. H. (2001), *Econometric Analysis of Panel Data*, Second Edition, John Wiley & Sons: Chichester, UK.

Baltagi, B. H. (1986), "An alternative Heteroscedastic Error Component Model, Problem", *Econometric Theory*, 4, 349-350.

Brégains, J. C., López, J. B. P. and Arcay, A. O. (2010), "Simulated Annealing Technique Applied to Mixed Logit Estimation", *Working paper*, Department of Electronic and Systems, Faculty of Informatics, La Coruna-Spain.
Browning, M. A. D. and Irish, M. (1985), "A profitable Approach to Labor Supply and Commodity Demands Over the Life Cycle", Econometrica, 53, 503-543.

Cameron, A. C. and Trivedi, P. K. (2005), "Repeated Cross Sections and Pseudo Panels", in Microeconometrics with Methods and Applications *published by the Cambridge University Press*, p.770.

Collado, M. D (1997), "Estimating Dynamic Models from Time Series of Independent Cross-Sections," *Journal of Econometrics*, 83, 37-62.

Collado, M. D (1998), "Estimating Binary Choice Models from Cohort Data", *Investigaciones Económicas*, 22, 259-276.

Deaton, A. (1986), "Panel Data from Time Series of Cross-Sections", *Journal of Econometrics*, 30, 109-126.

Durbin, J. (1954), "Errors in Variables," *Review of the International Statistical Institute*, 22, 23-32.

Dufour, J. M. and Stevanović (2013), "Factor-Augmented VARMA Models with Macroeconomic Applications", *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2013.818

Dufour, J. M. (2008), "Statistical Models and Likelihood Functions", *Working paper*, Department of Economics, McGill University.

Dufour, J. M. and Khalaf, L. (2000), "Simulation-Based Finite and Large Sample Tests in Multivariate Regressions" , *Cahier de recherche* No 2000-10, Département des sciences économiques, Université de Montréal.

Gardes, F., Duncan, G., Gaubert, P. and Starzec, C. (1999), "Panel and Pseudo Panel Estimation of Cross-Sections and Time Series Elasticities of Food Consumption: the case of American and Polish Data", *Cahiers de Recherche Team*, Université de Paris I.

Gardes, F., Langlois, S. and Richaudeau, D. (1996), "Cross Sections Versus Times Series Income Elasticities: an Estimation on a Pseudo Panel of Canadian Surveys", *Economics Letters*, 51, 169-175.

Greene, W. (2003), "Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model", *Working paper*, Department of Economics, Stern School of Business, New York University.

George, J. B. and Glenn, T. S. (1994), "A two-stage estimator for probit models with structural group effects", *Journal of Econometrics*, 64, 165-182.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, First Edition, Addison-Wesley Professional.

Gurgand, M., Gardes, F. and Bolduc, D.(1997), "Heteroscedasticity in Pseudo-Panel", *unpublished working paper, Lamia*, Université de Paris I.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

Hausman, J. A. and William, E. T. (1981), "Panel Data and Unobserved Individual Effects," *Econometrica*, 49, 1377-1398.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems*, University of Michigan Press, (Second edition: MIT Press. 1992).

Im, K. S., Seung, C. A., Peter, S. and Wooldridge, J. M. (1999), "Efficient estimation of Panel Data models with Strictly Exogenous Explanatory Variables," *Journal of Econo-*

metrics, 93, 177-201.

Kurkalova, L. A. and Rabotyagov, S. (2005), "Estimation of a Binary Choice Model with Grouped Choice Data", *Working paper 05-WP 381*, Center for Agricultural and Rural Development - Iowa State University.

Liu, Y. H. and Mahmassani, H. S. (1999), "Global maximum likelihood estimation procedure for multinomial probit (MNP) model parameters", *Elsevier Transportation Research Part B: Methodological*, 34, 419-449.
Mackinnon, J. G. (2002), "Bootstrap Inference in Econometrics", *The Canadian Journal of Economics/Revue canadienne d'Economique*, 615-645.

Mazodier, P. and Trognon, A. (1978), "Heteroscedasticity and Stratification in Error Components Models", *Annales de l'Insee*, vol. 30.31, 451-482.

McKenzie, David J. (2003), "Asymptotic Theory for Heterogeneous Dynamic Pseudo-panels," *Journal of Econometrics*, 120, 235-262.

Moffitt, R. (1993), "Identification and Estimation of Dynamic models with a Time Series of Repeated Cross-Sections," *Journal of Econometrics*, 59, 99-123.

Moffitt R. (1993), "Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections", *Journal of Econometrics*, 59, 99-123.

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

Paquet, M. F. et Bolduc, D. (2004), "Le problème des données longitudinales incomplètes: une nouvelle approche", *L'actualité économique, Erudit*, vol.80, 2-3, pp.341-361.

Pedro, A. D. G and Dean, F. H. (2007), "Initial Population for Genetic Algorithms: A Metric Approach", *Dicussion Paper*, School of Computer Science, University of Oklahoma.

Tamer, E. (2010), "Partial Identification in Econometrics", *The Annual Review of Economics*, 2:167-95.

Tsai, C. H., Leong, W., Mulley, C. and Clifton, G. (2012), "Examining estimator bias and efficiency for pseudo panel data: A Monte Carlo simulation approach", *Working paper ITLS-WP-12-07*, The University of Sydney.

Verbeek, M. (1995), "Pseudo-Panel Data," in László Mátyás and Patrick Sevestre ed. *The Econometrics of Panel data: A Handbook of the Theory with Applications*, Kluwer

Academic Publishers: Dordrecht, The Netherlands, pp.280-291.

Verbeek, M. and Theo, N. (1993), "Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-Sections," *Journal of Econometrics*, 59, 125-136.

Verbeek, M. and Francis, V. (2005), "Estimating Dynamic Models from Repeated Cross-Sections," *Journal of Econometrics*, 127, 83-102.

Verbeek, M. (1992), "Pseudo Panel Data", in Mátyás-Sevestre (eds), *The Econometrics of Panel Data*,chap.1, 303-315, Kluwer.

White, H. (1980), "A Heteroscedasticity Consistent Covariance Matrix Estimator and a direct Test for Heteroscedasticity", *Econometrica*, 48-4, 817-838.

Appendix

Simulations using Genetic Algorithm (GA)

$$\begin{cases} y^*_{i_{(t)}t} = x_{i_{(t)}t}\beta + \bar{x}_c\gamma + \xi_{i_{(t)}} + u_{i_{(t)}t} \\ y_{i_{(t)}t} = 1[x_{i_{(t)}t}\beta + \bar{x}_c\gamma + \xi_{i_{(t)}} + u_{i_{(t)}t} \geq 0] \\ \bar{y}_{ct} = \frac{1}{n_{ct}} \sum_{i=1, i\in c}^{n_{ct}} y_{i_{(t)}t} \end{cases}$$

Table 1: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $\xi_{i_{(t)}} \sim N(0,0.2)$, $u_{i_{(t)}t} \sim N(0,0.8)$.

	n_{ct}=10			n_{ct}=25			n_{ct}=50			n_{ct}=100		
	True			True			True			True		
$\hat{\beta}$ *bias*	1	−0.2097	−20.97	1	−0.1529	−15.29	1	−0.0650	−6.50	1	−0.0503	−5.03
MSE		0.0552	5.52		0.0333	3.33		0.0159	1.59		0.0140	1.40
$\hat{\gamma}$ *bias*	0.2	0.2174	108.70	0.2	0.0827	41.35	0.2	0.1350	67.50	0.2	0.0803	40.15
MSE		0.0615	153.75		0.0218	54.50		0.0360	90.00		0.0267	66.75
$\hat{m}_c$												
bias												
$C1$	0.3272	−0.0457	−13.96	0.3272	−0.0250	−7.64	0.3320	−0.0232	−6.98	0.3261	−0.0131	−4.02
$C2$	0.3199	−0.0417	−13.03	0.3225	−0.0164	−5.09	0.3226	−0.0123	−3.81	0.3235	−0.0074	−2.29
$C3$	0.3253	−0.0494	−15.18	0.3284	−0.0344	−10.47	0.3282	−0.0155	−4.72	0.3283	−0.0130	−3.96
$C4$	0.3511	−0.0571	−16.26	0.3433	−0.0526	−15.32	0.3377	−0.0234	−6.92	0.3350	−0.0152	−4.54
$C5$	0.3330	−0.0597	−17.92	0.3364	−0.0404	−12.00	0.3374	−0.0274	−8.12	0.3368	−0.0184	−5.46
$C6$	0.3257	−0.0527	−16.18	0.3408	−0.0481	−14.11	0.3397	−0.0282	−8.30	0.3363	−0.0193	−5.74
$C7$	0.3464	−0.0679	−19.60	0.3424	−0.0451	−13.17	0.3416	−0.0306	−8.95	0.3373	−0.0185	−5.48
$C8$	0.3506	−0.0686	−19.56	0.3402	−0.0462	−13.58	0.3351	−0.0227	−6.77	0.3405	−0.0258	−7.58
$C9$	0.3474	−0.0744	−21.41	0.3378	−0.0424	−12.55	0.3396	−0.0271	−7.98	0.3407	−0.0221	−6.49
$C10$	0.3469	−0.0845	−24.35	0.3488	−0.0530	−15.19	0.3529	−0.0388	−10.99	0.3468	−0.0286	−8.25
MSE												
$C1$		0.0027	2.52		0.0014	1.31		0.0014	1.27		0.0010	0.94
$C2$		0.0024	2.34		0.0011	1.06		0.0011	1.05		0.0010	0.95
$C3$		0.0031	2.92		0.0017	1.57		0.0010	0.92		0.0010	0.92
$C4$		0.0040	3.24		0.0033	2.80		0.0013	1.13		0.0010	0.89
$C5$		0.0042	3.78		0.0021	1.85		0.0014	1.22		0.0010	0.88
$C6$		0.0034	3.20		0.0028	2.41		0.0015	1.29		0.0011	0.97
$C7$		0.0052	4.33		0.0025	2.13		0.0017	1.45		0.0011	0.96
$C8$		0.0054	4.39		0.0026	2.24		0.0013	1.15		0.0012	1.03
$C9$		0.0061	5.05		0.0022	1.92		0.0014	1.21		0.0013	1.12
$C10$		0.0077	6.39		0.0036	2.95		0.0022	1.76		0.0015	1.24

Table 2: Bias and Relative Bias - Mean Squared Errors (MSE) and Relative Mean Squared Errors (RMSE) of the estimators: T=3, C=10, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $z_i \sim U(0,1)$, $\lambda_t \sim N(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $u_{i_{(t)}t} \sim N(0,0.8)$, $\xi_{i_{(t)}} \sim N(0,0.2)$.

		$n_{ct} = 100$			$n_{ct} = 200$		
		True			True		
$\hat{\beta}$	*bias*	1	-0.0503	-5.03	1	-0.0024	-0.24
	MSE		0.0140	1.40		0.0014	0.14
$\hat{\gamma}$	*bias*	0.2	0.0803	40.15	0.2	0.0080	4.00
	MSE		0.0267	66.75		0.0017	4.25

Table 3: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Errors (MSE) of the estimators: T=3, C=25, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $\xi_{i_{(t)}} \sim N(0,0.2)$, $u_{i_{(t)}t} \sim N(0,0.8)$.

		n_{ct}=10			n_{ct}=25			n_{ct}=50		
		True			True			True		
$\hat{\beta}$	*bias*	1	−0.1710	−17.10	1	−0.0905	−9.05	1	−0.0575	−5.75
	MSE		0.0406	4.06		0.0187	1.87		0.0156	1.56
$\hat{\gamma}$	*bias*	0.2	0.1689	84.45	0.2	0.1497	74.85	0.2	0.1300	65.00
	MSE		0.0456	114		0.0404	101		0.0356	89.00
$\hat{m}_c$										
	bias									
	$C1$	0.3272	−0.0252	−7.70	0.3272	−0.0028	−0.85	0.3320	−0.0016	−0.48
	$C2$	0.3199	−0.0167	−5.52	0.3225	−0.0011	−0.34	0.3226	−0.0061	−1.89
	$C3$	0.3253	−0.0381	−11.71	0.3284	−0.0031	−0.94	0.3282	−0.0010	−0.30
	$C4$	0.3511	−0.0539	−15.35	0.3433	−0.0171	−4.98	0.3377	−0.0024	−0.71
	$C5$	0.3330	−0.0404	−12.13	0.3364	−0.0089	−2.64	0.3374	−0.0024	−0.71
	$C6$	0.3257	−0.0230	−7.06	0.3408	−0.0144	−4.22	0.3397	−0.0030	−0.88
	$C7$	0.3464	−0.0426	−12.29	0.3424	−0.0166	−4.84	0.3416	−0.0031	−0.90
	$C8$	0.3506	−0.0629	−17.94	0.3402	−0.0141	−4.14	0.3351	−0.0024	−0.71
	$C9$	0.3474	−0.0494	−14.21	0.3378	−0.0143	−4.23	0.3396	−0.0024	−0.70
	$C10$	0.3469	−0.0537	−15.48	0.3488	−0.0232	−6.65	0.3529	−0.0037	−1.04
	$C11$	0.3446	−0.0427	−12.39	0.3461	−0.0230	−6.64	0.3394	−0.0027	−0.79
	$C12$	0.3577	−0.0548	−15.32	0.3513	−0.0275	−7.82	0.3421	−0.0028	−0.81
	$C13$	0.3437	−0.0567	−16.49	0.3468	−0.0229	−6.60	0.3471	−0.0033	−0.95
	$C14$	0.3535	−0.0564	−15.95	0.3474	−0.0235	−6.76	0.3488	−0.0038	−1.08
	$C15$	0.3484	−0.0559	−16.04	0.3493	−0.0230	−6.58	0.3463	−0.0034	−0.98
	$C16$	0.3523	−0.0495	−14.05	0.3458	−0.0177	−5.11	0.3432	−0.0030	−0.87
	$C17$	0.3632	−0.0593	−16.32	0.3524	−0.0236	−6.69	0.3509	−0.0036	−1.02
	$C18$	0.3572	−0.0693	−19.40	0.3501	−0.0213	−6.08	0.3493	−0.0035	−1.00
	$C19$	0.3517	−0.0536	−15.24	0.3425	−0.0154	−4.49	0.3494	−0.0030	−0.85
	$C20$	0.3455	−0.0521	−15.07	0.3458	−0.0202	−5.84	0.3476	−0.0033	−0.94

	n_{ct}=10			n_{ct}=25			n_{ct}=50		
	True	Est	REst(%)	True	Est	REst(%)	True	Est	REst(%)
$C21$	0.3600	−0.0580	−16.11	0.3515	−0.0266	−7.56	0.3428	−0.0027	−0.78
$C22$	0.3549	−0.0519	−14.62	0.3498	−0.0242	−6.91	0.3447	−0.0030	−0.87
$C23$	0.3491	−0.0621	−17.78	0.3462	−0.0194	−5.60	0.3473	−0.0034	−0.97
$C24$	0.3490	−0.0516	−14.78	0.3413	−0.0149	−4.36	0.3442	−0.0029	−0.84
$C25$	0.3570	−0.0644	−18.03	0.3484	−0.0241	−6.91	0.3467	−0.0031	−0.89
MSE									
$C1$		0.0014	1.30		0.0013	1.21		0.0012	1.08
$C2$		0.0011	1.07		0.0016	1.53		0.0010	0.96
$C3$		0.0022	2.07		0.0015	1.39		0.0011	1.02
$C4$		0.0037	3.00		0.0018	1.52		0.0014	1.22
$C5$		0.0024	2.16		0.0015	1.32		0.0015	1.31
$C6$		0.0013	1.22		0.0016	1.37		0.0022	1.90
$C7$		0.0026	2.16		0.0016	1.36		0.0023	1.97
$C8$		0.0046	3.74		0.0016	1.38		0.0015	1.33
$C9$		0.0032	2.65		0.0017	1.48		0.0017	1.47
$C10$		0.0036	2.99		0.0019	1.56		0.0020	1.60
$C11$		0.0026	2.18		0.0021	1.75		0.0020	1.73
$C12$		0.0038	2.96		0.0023	1.86		0.0018	1.53
$C13$		0.0040	3.38		0.0021	1.74		0.0020	1.66
$C14$		0.0040	3.20		0.0022	1.82		0.0028	2.30
$C15$		0.0039	3.21		0.0021	1.72		0.0020	1.66
$C16$		0.0032	2.57		0.0020	1.67		0.0020	1.69
$C17$		0.0043	3.25		0.0023	1.85		0.0020	1.62
$C18$		0.0055	4.31		0.0019	1.55		0.0023	1.88
$C19$		0.0037	2.99		0.0017	1.44		0.0016	1.31
$C20$		0.0035	2.93		0.0020	1.67		0.0020	1.65
$C21$		0.0042	3.24		0.0024	1.94		0.0020	1.70
$C22$		0.0035	2.77		0.0022	1.79		0.0021	1.76
$C23$		0.0046	3.77		0.0019	1.58		0.0019	1.57
$C24$		0.0035	2.87		0.0016	1.37		0.0014	1.18
$C25$		0.0049	3.84		0.0020	1.64		0.0020	1.66

Table 4: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Error (RMSE) of the estimators: T=3, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $\xi_{i_{(t)}} \sim N(0,0.2)$, $u_{i_{(t)}t} \sim N(0,0.8)$.

	C=10, n_{ct}=25			C=25,n_{ct}=10			C=10, n_{ct}=50			C=50, n_{ct}=10		
	True			True			True			True		
$\hat{\beta}$ *bias*	1	−0.1529	−15.29	1	−0.1710	−17.10	1	−0.0650	−6.50	1	−0.1091	−10.91
MSE		0.0333	3.33		0.0406	4.06		0.0159	1.59		0.0390	3.90
$\hat{\gamma}$ *bias*	0.2	0.0827	41.35	0.2	0.1689	84.45	0.2	0.1350	67.50	0.2	0.1491	74.55
MSE		0.0218	54.50		0.0456	114		0.0360	90.00		0.0410	102.50
$\hat{m}_c$												
bias												
$C1$	0.3272	−0.0250	−7.64	0.3272	−0.0252	−7.70	0.3320	−0.0232	−6.98	0.3301	−0.0242	−7.33
$C2$	0.3225	−0.0164	−5.09	0.3199	−0.0167	−5.22	0.3226	−0.0123	−3.81	0.3264	−0.0136	−4.16
$C3$	0.3284	−0.0344	−10.47	0.3253	−0.0381	−11.71	0.3282	−0.0155	−4.72	0.3378	−0.0333	−9.85
$C4$	0.3433	−0.0526	−15.32	0.3511	−0.0539	−15.35	0.3377	−0.0234	−6.92	0.3644	−0.0434	−11.91
$C5$	0.3364	−0.0404	−12.00	0.3330	−0.0404	−12.13	0.3374	−0.0274	−8.12	0.3473	−0.0315	−9.07
$C6$	0.3408	−0.0481	−14.11	0.3257	−0.0230	−7.06	0.3397	−0.0282	−8.30	0.3382	−0.0245	−7.24
$C7$	0.3424	−0.0451	−13.17	0.3464	−0.0426	−12.29	0.3416	−0.0306	−8.95	0.3532	−0.0406	−11.49
$C8$	0.3402	−0.0462	−13.58	0.3506	−0.0629	−17.94	0.3351	−0.0227	−6.77	0.3499	−0.0401	−11.46
$C9$	0.3378	−0.0424	−12.55	0.3474	−0.0494	−14.21	0.3396	−0.0271	−7.98	0.3533	−0.0353	−9.99
$C10$	0.3488	−0.0530	−15.19	0.3469	−0.0537	−15.48	0.3529	−0.0388	−10.99	0.3508	−0.0482	−13.70
MSE												
$C1$		0.0014	1.31		0.0014	1.30		0.0014	1.27		0.0014	1.28
$C2$		0.0011	1.06		0.0011	1.07		0.0011	1.05		0.0011	1.03
$C3$		0.0017	1.57		0.0022	2.07		0.0010	0.92		0.0015	1.31
$C4$		0.0033	2.80		0.0037	3.00		0.0013	1.13		0.0030	2.25
$C5$		0.0021	1.85		0.0024	2.16		0.0014	1.22		0.0017	1.40
$C6$		0.0028	2.41		0.0013	1.22		0.0015	1.29		0.0015	1.31
$C7$		0.0025	2.13		0.0026	2.16		0.0017	1.45		0.0024	1.92
$C8$		0.0026	2.24		0.0046	3.74		0.0013	1.15		0.0025	2.04
$C9$		0.0022	1.92		0.0032	2.65		0.0014	1.21		0.0020	1.60
$C10$		0.0036	2.95		0.0036	2.99		0.0022	1.76		0.0032	2.60

Table 5: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $n_{ct} = 100$, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$.

		$u_{i_{(t)}t} \sim N(0,0.8)$, $\xi_{i_{(t)}} \sim N(0,0.2)$			$u_{i_{(t)}t} \sim logistic(0, \frac{0.2432\pi^2}{3})$, $\xi_{i_{(t)}} \sim N(0,0.2)$		
		True	Est	REst(%)	True	Est	REst(%)
$\hat{\beta}$	$bias$	1	-0.0503	-5.03	1	-0.2416	-24.16
	MSE		0.0140	1.40		0.0693	6.93
$\hat{\gamma}$	$bias$	0.2	0.0803	40.15	0.2	0.1585	79.25
	MSE		0.0267	66.75		0.0407	101.75
$\hat{m}_c$							
	$bias$						
	$C1$	0.3261	-0.0131	-4.02	0.3261	-0.0658	-20.17
	$C2$	0.3235	-0.0074	-2.29	0.3235	-0.0614	-18.97
	$C3$	0.3283	-0.0130	-3.96	0.3283	-0.0667	-20.31
	$C4$	0.3350	-0.0152	-4.54	0.3350	-0.0706	-21.07
	$C5$	0.3368	-0.0184	-5.46	0.3368	-0.0732	-21.73
	$C6$	0.3363	-0.0193	-5.74	0.3363	-0.0736	-21.88
	$C7$	0.3373	-0.0185	-5.48	0.3373	-0.0735	-21.79
	$C8$	0.3405	-0.0258	-7.58	0.3405	-0.0792	-23.25
	$C9$	0.3407	-0.0221	-6.49	0.3407	-0.0771	-22.62
	$C10$	0.3468	-0.0286	-8.25	0.3468	-0.0834	-24.04
	MSE						
	$C1$		0.0010	0.94		0.0047	4.41
	$C2$		0.0010	0.95		0.0041	3.91
	$C3$		0.0010	0.92		0.0048	4.45
	$C4$		0.0010	0.89		0.0053	4.72
	$C5$		0.0010	0.88		0.0056	4.93
	$C6$		0.0011	0.97		0.0057	5.03
	$C7$		0.0011	0.96		0.0057	5.01
	$C8$		0.0012	1.03		0.0066	5.69
	$C9$		0.0013	1.12		0.0063	5.42
	$C10$		0.0015	1.24		0.0072	5.98

Table 6: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $n_{ct} = 100$, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $u_{i_{(t)}t} \sim N(0,0.8)$, $\xi_{i_{(t)}} \sim N(0,0.2)$.

		$\eta_{i_{(t)}t} \sim U(0,1)$			$\eta_{i_{(t)}t} \sim N(0,1)$		
		True	Est	REst(%)	True	Est	REst(%)
$\hat{\beta}$	$bias$	1	-0.0503	-5.03	1	-0.0001	-0.01
	MSE		0.0140	1.40		0.0011	0.11
$\hat{\gamma}$	$bias$	0.2	0.0803	40.15	0.2	-0.0461	-23.05
	MSE		0.0267	66.75		0.0262	65.50
$\hat{m}_c$							
	$bias$						
	$C1$	0.3261	-0.0131	-4.02	0.3261	-0.0072	-2.50
	$C2$	0.3235	-0.0074	-2.29	0.3235	0.0001	0.03
	$C3$	0.3283	-0.0130	-3.96	0.3283	-0.0012	-0.43
	$C4$	0.3350	-0.0152	-4.54	0.3350	-0.0008	-0.28
	$C5$	0.3368	-0.0184	-5.46	0.3368	-0.0025	-0.88
	$C6$	0.3363	-0.0193	-5.74	0.3363	-0.0034	-1.24
	$C7$	0.3373	-0.0185	-5.48	0.3373	-0.0035	-1.23
	$C8$	0.3405	-0.0258	-7.58	0.3405	0.0048	1.76
	$C9$	0.3407	-0.0221	-6.49	0.3407	0.0008	0.28
	$C10$	0.3468	-0.0286	-8.25	0.3468	0.0076	2.77
	MSE						
	$C1$		0.0010	0.94		0.0008	0.24
	$C2$		0.0010	0.95		0.0009	0.28
	$C3$		0.0010	0.92		0.0009	0.27
	$C4$		0.0010	0.89		0.0008	0.23
	$C5$		0.0010	0.88		0.0007	0.20
	$C6$		0.0011	0.97		0.0008	0.23
	$C7$		0.0011	0.96		0.0006	0.17
	$C8$		0.0012	1.03		0.0009	0.26
	$C9$		0.0013	1.12		0.0006	0.17
	$C10$		0.0015	1.24		0.0009	0.25

Table 7: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $n_{ct} = 100$, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $u_{i_{(t)}t} \sim N(0,0.8)$, $\xi_{i_{(t)}} \sim N(0,0.2)$.

		$\lambda_t \sim N(0,1)$			$\lambda_t \sim N(0,10)$		
		True			True		
$\hat{\beta}$	*bias*	1	−0.0503	−5.03	1	−0.0093	−0.93
	MSE		0.0140	1.40		0.0025	0.25
$\hat{\gamma}$	*bias*	0.2	0.0803	40.15	0.2	0.0111	5.50
	MSE		0.0267	66.75		0.0036	9.00
$\hat{m}_c$							
	bias						
	$C1$	0.3261	−0.0131	−4.02	0.3287	−0.0022	−0.66
	$C2$	0.3235	−0.0074	−2.29	0.3305	−0.0024	−0.72
	$C3$	0.3283	−0.0130	−3.96	0.3384	−0.0030	−0.88
	$C4$	0.3350	−0.0152	−4.54	0.3464	−0.0034	−0.98
	$C5$	0.3368	−0.0184	−5.46	0.3481	−0.0039	−1.12
	$C6$	0.3363	−0.0193	−5.74	0.3465	−0.0039	−1.12
	$C7$	0.3373	−0.0185	−5.48	0.3451	−0.0035	−1.01
	$C8$	0.3405	−0.0258	−7.58	0.3449	−0.0036	−1.04
	$C9$	0.3407	−0.0221	−6.49	0.3355	−0.0027	−0.80
	$C10$	0.3468	−0.0286	−8.25	0.3325	−0.0022	−0.66
	MSE						
	$C1$		0.0010	0.94		0.0001	0.09
	$C2$		0.0010	0.95		0.0001	0.09
	$C3$		0.0010	0.92		0.0002	0.17
	$C4$		0.0010	0.89		0.0002	0.16
	$C5$		0.0010	0.88		0.0002	0.16
	$C6$		0.0011	0.97		0.0002	0.16
	$C7$		0.0011	0.96		0.0002	0.16
	$C8$		0.0012	1.03		0.0002	0.16
	$C9$		0.0013	1.12		0.0001	0.08
	$C10$		0.0015	1.24		0.0001	0.09

Table 8: Bias and Relative Bias - Mean Squared Error and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $n_{ct} = 100$, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $\xi_{i_{(t)}} \sim N(0,0.2)$, $u_{i_{(t)}t} \sim N(0,0.8)$.

	$\beta = -1 \quad \gamma = 0.2$			$\beta = 0.5 \quad \gamma = 0.2$			$\beta = 1 \quad \gamma = 0.2$			$\beta = 1.5 \quad \gamma = 0.2$		
	True			True			True			True		
$\hat{\beta}$ *bias*	-1	0.0421	-4.21	0.5	-0.0145	-2.90	1	-0.0503	-5.03	1.5	-0.0210	-1.40
MSE		0.0394	3.94		0.0099	3.96		0.0140	1.40		0.0554	2.46
$\hat{\gamma}$ *bias*	0.2	-0.0642	-32.10	0.2	0.0547	27.35	0.2	0.0803	40.15	0.2	0.0620	31.00
MSE		0.0240	60.00		0.0230	57.50		0.0267	66.75		0.0211	52.75
$\hat{m}_c$												
bias												
$C1$	-0.3539	0.0147	-4.15	0.1868	-0.0069	-3.69	0.3261	-0.0131	-4.02	0.4077	-0.0145	-3.55
$C2$	-0.3538	0.0144	-4.07	0.1863	-0.0081	-4.34	0.3235	-0.0074	-2.29	0.4025	-0.0125	-3.10
$C3$	-0.3562	0.0147	-4.12	0.1673	-0.0069	-4.12	0.3283	-0.0130	-3.96	0.4133	-0.0152	-3.67
$C4$	-0.3599	0.0148	-4.11	0.1887	-0.0065	-3.44	0.3350	-0.0152	-4.54	0.4286	-0.0211	-4.92
$C5$	-0.3612	0.0151	-4.18	0.1890	-0.0059	-3.12	0.3368	-0.0184	-5.46	0.4329	-0.0284	-6.56
$C6$	-0.3600	0.0150	-4.16	0.1889	-0.0057	-3.01	0.3363	-0.0193	-5.74	0.4321	-0.0302	-6.98
$C7$	-0.3605	0.0149	-4.13	0.1891	-0.0059	-3.12	0.3373	-0.0185	-5.48	0.4350	-0.0292	-6.71
$C8$	-0.3623	0.0153	-4.22	0.1897	-0.0044	-2.31	0.3405	-0.0258	-7.58	0.4426	-0.0361	-8.15
$C9$	-0.3622	0.0151	-4.16	0.1897	-0.0052	-2.74	0.3407	-0.0221	-6.49	0.4441	-0.0291	-6.55
$C10$	-0.3655	0.0155	-4.24	0.1910	-0.0039	-2.04	0.3468	-0.0286	-8.25	0.4573	-0.0330	-7.21
MSE												
$C1$		0.0022	1.75		0.00008	0.23		0.0010	0.94		0.0026	1.56
$C2$		0.0020	1.59		0.00009	0.26		0.0010	0.95		0.0015	0.92
$C3$		0.0021	1.65		0.00008	0.28		0.0010	0.92		0.0027	1.58
$C4$		0.0022	1.69		0.00007	0.19		0.0010	0.89		0.0026	1.41
$C5$		0.0022	1.68		0.00006	0.17		0.0010	0.88		0.0030	1.60
$C6$		0.0022	1.69		0.00006	0.17		0.0011	0.97		0.0030	1.60
$C7$		0.0022	1.69		0.00007	0.20		0.0011	0.96		0.0030	1.58
$C8$		0.0023	1.75		0.00006	0.17		0.0012	1.03		0.0040	2.04
$C9$		0.0023	1.75		0.00006	0.17		0.0013	1.12		0.0033	1.67
$C10$		0.0024	1.79		0.00005	0.14		0.0015	1.24		0.0038	1.81

Table 9: Bias and Relative Bias - Mean Squared Error and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $n_{ct} = 100$, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $\xi_{i_{(t)}} \sim N(0,0.2)$, $u_{i_{(t)}t} \sim N(0,0.8)$.

	$\beta = 1 \quad \gamma = -0.7$			$\beta = 1 \quad \gamma = 0$			$\beta = 1 \quad \gamma = 0.2$			$\beta = 1 \quad \gamma = 0.7$		
	True			True			True			True		
$\hat{\beta}$ $bias$	1	−0.0350	−3.50	1	−0.0086	−0.86	1	−0.0503	−5.03	1	−0.0533	−5.33
MSE		0.017	1.70		0.0026	0.26		0.0140	1.40		0.0139	1.39
$\hat{\gamma}$ $bias$	−0.7	0.290	−41.43	0	0.0163	...	0.2	0.0803	40.15	0.7	0.2730	39.00
MSE		0.3254	66.41		0.0260	...		0.0267	66.75		0.1450	29.59
$\hat{m}_c$												
$bias$												
$C1$	0.3772	−0.0130	−3.44	0.3423	−0.00301	−0.87	0.3261	−0.0131	−4.02	0.2776	0.0040	1.44
$C2$	0.3755	−0.0128	−3.40	0.3399	−0.00251	−0.73	0.3235	−0.0074	−2.29	0.2747	0.0063	2.29
$C3$	0.3759	−0.0129	−3.43	0.3434	−0.00293	−0.85	0.3283	−0.0130	−3.96	0.2828	−0.0028	−0.99
$C4$	0.3772	−0.0129	−3.41	0.3485	−0.00305	−0.87	0.3350	−0.0152	−4.54	0.2936	−0.0087	−2.96
$C5$	0.3783	−0.0130	−3.43	0.3501	−0.00333	−0.95	0.3368	−0.0184	−5.46	0.2964	−0.0129	−4.35
$C6$	0.3765	−0.0129	−3.42	0.3492	−0.00336	−0.96	0.3363	−0.0193	−5.74	0.2971	−0.0151	−5.08
$C7$	0.3767	−0.0129	−3.42	0.3499	−0.00387	−1.09	0.3405	−0.0258	−7.58	0.3042	−0.0254	−8.34
$C9$	0.3773	−0.0129	−3.41	0.3524	−0.00354	−1.00	0.3407	−0.0221	−6.49	0.3055	−0.0221	−7.23
$C10$	0.3785	−0.0131	−3.46	0.3570	−0.00405	−1.13	0.3468	−0.0286	−8.25	0.3156	−0.0223	−7.06
MSE												
$C1$		0.0017	1.19		0.00015	0.12		0.0010	0.94		0.0015	1.94
$C2$		0.0016	1.13		0.00011	0.09		0.0010	0.95		0.0013	1.72
$C3$		0.0016	1.13		0.00014	0.11		0.0010	0.92		0.0015	1.87
$C4$		0.0016	1.12		0.00014	0.11		0.0010	0.89		0.0014	1.62
$C5$		0.0017	1.18		0.00015	0.12		0.0010	0.88		0.0013	1.47
$C6$		0.0016	1.12		0.00016	0.13		0.0011	0.97		0.0017	1.92
$C7$		0.0016	1.12		0.00016	0.13		0.0011	0.96		0.0017	1.90
$C8$		0.0017	1.19		0.00020	0.16		0.0012	1.03		0.0018	1.94
$C9$		0.0016	1.12		0.00018	0.14		0.0013	1.12		0.0018	1.92
$C10$		0.0017	1.18		0.00021	0.16		0.0015	1.24		0.0019	1.90

Table 10: Bias and Relative Bias - Mean Squared Error and Relative Mean Squared Error (RMSE) of the estimators: C=10, $n_{ct} = 100$, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $u_{i_{(t)}t} \sim N(0,0.8)$, $\xi_{i_{(t)}} \sim N(0,0.2)$.

		$T=3$			$T=5$		
		True			True		
$\hat{\beta}$	*bias*	1	-0.0503	-5.03	1	-0.0106	-1.06
	MSE		0.0140	1.40		0.0283	2.83
$\hat{\gamma}$	*bias*	0.2	0.0803	40.15	0.2	0.0142	7.10
	MSE		0.0267	66.75		0.0318	79.50
$\hat{m}_c$							
	bias						
	$C1$	0.3261	-0.0131	-4.02	0.3259	-0.0138	-4.23
	$C2$	0.3235	-0.0074	-2.29	0.3282	-0.0154	-4.69
	$C3$	0.3283	-0.0130	-3.96	0.3276	-0.0194	-5.92
	$C4$	0.3350	-0.0152	-4.54	0.3235	-0.0133	-4.11
	$C5$	0.3368	-0.0184	-5.46	0.3252	-0.0158	-4.85
	$C6$	0.3363	-0.0193	-5.74	0.3273	-0.0171	-5.22
	$C7$	0.3373	-0.0185	-5.48	0.3269	-0.0164	-5.01
	$C8$	0.3405	-0.0258	-7.58	0.3257	-0.0142	-4.35
	$C9$	0.3407	-0.0221	-6.49	0.3261	-0.0163	-4.99
	$C10$	0.3468	-0.0286	-8.25	0.3243	-0.0164	-5.05
	MSE						
	$C1$		0.0010	0.94		0.0009	0.84
	$C2$		0.0010	0.95		0.0010	0.92
	$C3$		0.0010	0.92		0.0009	0.83
	$C4$		0.0010	0.89		0.0010	0.95
	$C5$		0.0010	0.88		0.0012	1.13
	$C6$		0.0011	0.97		0.0012	1.12
	$C7$		0.0011	0.96		0.0010	0.93
	$C8$		0.0012	1.03		0.0010	0.94
	$C9$		0.0013	1.12		0.0010	0.94
	$C10$		0.0015	1.24		0.0013	1.23

Table 11: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $\lambda_t \sim N(0,1)$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $\xi_{i_{(t)}} \sim N(0,0.2)$, $u_{i_{(t)}t} \sim N(0,0.8)$.

	$\beta = 1$ $\gamma = 0.2$			$\beta = 1$ $\gamma = unheeded$			$\beta = unheeded$ $\gamma = 0.2$		
	True			True			True		
$\hat{\beta}$ *bias*	1	-0.0503	- 5.03	1	-0.1736	-17.36	1	...	...
MSE		0.0140	1.40		0.0430	4.30		...	...
$\hat{\gamma}$ *bias*	0.2	0.0803	40.15	0.2	...	...	0.2	0.9504	475.20
MSE		0.0267	66.75		...	...		0.0917	229.25
$\hat{m}_c$									
bias									
$C1$	0.3261	-0.0131	- 4.02	0.3423	-0.0410	- 11.97	0.0007939	0.00032	40.307
$C2$	0.3235	-0.0074	-2.29	0.3399	-0.0479	- 14.09	0.0007938	0.00032	40.312
$C3$	0.3283	-0.0130	- 3.96	0.3434	-0.0432	- 12.58	0.0007941	0.00032	40.297
$C4$	0.3350	-0.0152	- 4.54	0.3485	-0.0448	-12.85	0.0007945	0.00032	40.276
$C5$	0.3368	-0.0184	-5.46	0.3501	-0.0411	- 11.73	0.0007946	0.00032	40.271
$C6$	0.3363	-0.0193	- 5.74	0.3492	-0.0399	- 11.42	0.0007947	0.00032	40.266
$C7$	0.3373	-0.0185	- 5.48	0.3499	-0.0419	- 11.97	0.0007947	0.00032	40.266
$C8$	0.3405	-0.0258	- 7.58	0.3524	-0.0336	- 9.53	0.0007949	0.00032	40.256
$C9$	0.3407	-0.0221	- 6.49	0.3524	-0.0393	- 11.15	0.0007949	0.00032	40.256
$C10$	0.3468	-0.0286	- 8.25	0.3570	-0.0337	-9.43	0.0007953	0.00032	40.236
MSE									
$C1$		0.0010	0.94		0.0031	2.64		0.000001	158.66
$C2$		0.0010	0.95		0.0036	3.11		0.000001	158.70
$C3$		0.0010	0.92		0.0034	2.88		0.000001	158.58
$C4$		0.0010	0.89		0.0032	2.63		0.000001	158.42
$C5$		0.0010	0.88		0.0026	2.12		0.000001	158.38
$C6$		0.0011	0.97		0.0029	2.37		0.000001	158.34
$C7$		0.0011	0.96		0.0031	2.53		0.000001	158.34
$C8$		0.0012	1.03		0.0024	1.93		0.000001	158.26
$C9$		0.0013	1.12		0.0029	2.33		0.000001	158.26
$C10$		0.0015	1.24		0.0023	1.80		0.000001	158.10

Table 12: Bias and Relative Bias - Mean Squared Error (MSE) and Relative Mean Squared Error (RMSE) of the estimators: T=3, C=10, $n_{ct} = 100$, $x_{i_{(t)}t} = \lambda_t z_i + \eta_{i_{(t)}t}$, $z_i \sim U(0,1)$, $\eta_{i_{(t)}t} \sim U(0,1)$, $u_{i_{(t)}t} \sim N(0,0.8)$, $\xi_{i_{(t)}} \sim N(0,0.2)$, $\lambda_t \sim N(0,10)$.

		Only Grouped Choices Observed			Individual Choices observed		
		True			True		
$\hat{\beta}$	*bias*	1	-0.0093	-0.93	1	-0.0094	-0.94
	MSE		0.0025	0.25		0.0013	0.13
$\hat{\gamma}$	*bias*	0.2	0.0111	5.50	0.2	0.0101	5.05
	MSE		0.0036	9.00		0.0020	5.00
$\hat{m}_c$							
	bias						
	$C1$	0.3287	-0.0022	-0.66	0.3287	-0.0022	-0.66
	$C2$	0.3305	-0.0024	-0.72	0.3305	-0.0025	-0.76
	$C3$	0.3384	-0.0030	-0.88	0.3384	-0.0027	-0.80
	$C4$	0.3464	-0.0034	-0.98	0.3464	-0.0036	-1.03
	$C5$	0.3481	-0.0039	-1.12	0.3481	-0.0037	-1.06
	$C6$	0.3465	-0.0039	-1.12	0.3465	-0.0038	-1.10
	$C7$	0.3451	-0.0035	-1.01	0.3451	-0.0037	-1.07
	$C8$	0.3449	-0.0036	-1.04	0.3449	-0.0037	-1.07
	$C9$	0.3355	-0.0027	-0.80	0.3355	-0.0027	-0.80
	$C10$	0.3325	-0.0022	-0.66	0.3325	-0.0023	-0.69
	MSE						
	$C1$		0.0001	0.09		0.00002	0.0185
	$C2$		0.0001	0.09		0.00002	0.0183
	$C3$		0.0002	0.17		0.00003	0.0262
	$C4$		0.0002	0.16		0.00002	0.0167
	$C5$		0.0002	0.16		0.00003	0.0248
	$C6$		0.0002	0.16		0.00003	0.0250
	$C7$		0.0002	0.16		0.00003	0.0252
	$C8$		0.0002	0.16		0.00002	0.0168
	$C9$		0.0001	0.08		0.00002	0.0178
	$C10$		0.0001	0.09		0.00002	0.0181