UNIVERSITÀ POLITECNICA DELLE MARCHE
Facoltà di Economia G.Fuà

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI (DISES)
Ph.D. degree in Economics - XXXII cycle

Doctoral Thesis

Three Essays on the Conditional Inference Approach for Binary Panel Data Models

Author:
Francesco Valentini

Supervisors:
Prof. Francesco Bartolucci
Dr. Claudia Pigini

Contents

List of Tables

List of Figures

Ringraziamenti

Come al solito sarò di poche parole. Sono stati anni intensi, belli e di duro lavoro. Approfitto dell'occasione per ringraziare le tante persone che mi sono state vicine e non posso che iniziare dalla mia famiglia; grazie a loro ed al continuo sostegno è stato tutto molto più facile.

La mia più sincera gratitudine va ai miei Supervisors che con grandissima professionalità ed umanità mi hanno accompagnato in questi mesi. A loro devo tanto, la loro capacità di trasmettere conoscenza e passione è stata fondamentale nel mio percorso.

Ho anche avuto la fortuna di condividere questi mesi con dei veri e propri amici, non solo dei colleghi. Anche a loro un caloroso ringraziamento, così come a tutto il Dipartimento, in cui mi sono sempre sentito a casa.

Una menzione particolare va a Chiara e Luca, il nostro legame va ben oltre una semplice amicizia e raramente nella vita capita di avere compagni di viaggio così speciali.

Grazie a tutti,
Francesco

Introduction

The use of binary panel data models with fixed-effects has become relevant in many empirical investigations of individual behaviour and, in general, in microeconometric analysis. Moreover, the interest in the fixed-effects approach is due to its lack of parametric assumptions concerning unobserved heterogeneity. The key issue underlying the parameters estimation of these models is the *incidental parameters problem* which makes the Maximum Likelihood Estimator inconsistent. The literature has been growing rapidly in the recent decades and many different techniques aimed at overcome such problem have been developed. Conditional inference is a viable approach since it exploits sufficient statistics in order to get rid of nuisance parameters, providing consistent estimators for the parameters of interest for the Logit model.

Even if attractive in theory, the application of the conditional inference framework is hampered by the lack of some extensions to deal with real data. In particular, the aim of the present work is trying to overcome two different limitations of this approach: (i) the issue of endogenous data selection mechanisms, and (ii) the computational burden of the conditional likelihood functions of these models.

Endogenous data selection mechanisms consist in dealing with two different problems: an endogenous binary explanatory variable and the sample selection issue. Both these aspects could lead to inconsistent estimators and it is crucial to test whether the selection is endogenous. In this regard, I propose a simple procedure that allows practitioners to test for these forms of endogeneity in a pure fixed-effects approach. The impossibility of jointly modeling the main outcome and the selection variable within the Logit model framework is handled by an approximating model which is estimated by a

Pseudo Conditional Maximum Likelihood (PCML) estimator. The PCML estimator is then exploited in order to perform the test. Furthermore, finite sample performance of the test is evaluated by a Monte Carlo simulation showing a good performance in terms of size and power. Finally, an empirical illustration to real data concerning the relationship between health status and retirement, based on the the Survey of Health, Ageing and Retirement in Europe, is provided.

The issue concerning the computational burden of the conditional likelihood functions considered in this work is closely related to the number of time occasions in the panel dataset. As a matter of fact, the computation of the likelihood function is not feasible for large time dimensions, limiting the applicability of the conditional inference framework. Computational problems of the static conditional logit model are dealt with a recursive algorithm while the extension of the recursive computation for dynamic models is an unexplored field. Here, I propose a recursive algorithm for the computation of the conditional likelihood function of a class of dynamic models, namely, the Quadratic Exponential (QE) model and its extensions. The proposed algorithm allows these models to be estimated also for large time dimensions, where the alternative algebric computation would have been infeasible. As an example, an application to real data concerning brand loyalty is proposed, where the QE model parameters are estimated exploiting the proposed recursion.

The work is organised as follows: Chapter 1 reviews the literature concerning the estimation of binary panel data models and an extensive simulation study that aims to valuate the performance of some recently proposed estimators for the Dynamic Logit model parameters. Chapter 2 shows the proposed methodology to handle the issue of testing for endogenous selection mechanisms and, finally, Chapter 3 presents the proposed recursive algorithm for the QE models.

Chapter 1

Estimation of nonlinear binary panel data models

1.1 Introduction

Panel data analysis plays a major role in Econometrics. The related literature has been rapidly growing during the last decades and it is of main interest for researchers and practitioners. A panel data set, differently from cross-sectional or time series data, provides repeated observations over time for every unit in the sample. This kind of data structure is more informative than cross-sectional data but, at the same time, technical issues arise. One key point is the unobserved heterogeneity across units and over time. The behaviour of an individual is influenced by characteristics that cannot be directly observed and controlled for by the analyst, for example the risk attitude for investments or the personal preferences about consumption and saving. Therefore, accounting for these factors becomes crucial in the formulation of econometric models and panel data allows the analysts to control for some unobserved components.

Consider a sample of n individuals observed for T time periods, a continuous dependent variable y_{it}, where the subscript it denotes the observation for the i-th individual at the t-th period, and a set of exogenous covariates collected in a column vector of dimension $k \times 1$, denoted by $\boldsymbol{x}_{it}$. A general specification is a linear model based on the assumption

$$y_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it},$$

where α_i denotes the individual unobserved heterogeneity that is represented by a unit-specific intercept. The role played by α_i is to capture time-invariant unobservable characteristics for the i-th individual in the sample that cannot be described by the covariates. It is also possible to take into account heterogeneity over time including time dummies in $\boldsymbol{x}_{it}$. Moreover, $\boldsymbol{\beta}$ is a vector of regression parameters and ε_{it} is the error component, which is assumed to be a random variable with 0 mean and constant finite variance σ^2_ε, representing idiosyncratic shocks. Hsiao (2014) provides a wide analysis concerning estimation and inferential procedures for panel data models.

While linear models play an important role in econometric literature, they show some drawbacks when the domain of the response variable is restricted such as for binary and discrete outcomes. In these cases, the Generalized Linear Model (GLM) formulation (McCullagh and Nelder, 1989) allows for a higher degree of flexibility. GLMs are based on three components: (i) a distribution belonging to the exponential family, such as Poisson, Binomial or Gamma for the response variable y_{it}; (ii) a systematic component which is the linear predictor $\eta_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}$; (iii) the link function $g(\cdot)$ that specifies the relationship between the expected value of the response variable, $\mathrm{E}(y_{it}) = \mu_{it}$, and the linear predictor η_{it}, such that $\eta_{it} = g(\mu_{it})$. The linear model is a peculiar case given by the identity link function where the expected value of y_{it} coincides with the linear predictor, that is, $\eta_{it} = \mu_{it}$. Hence, the expected value of y_{it} may potentially assume any value in $\mathbb{R}$. A prominent case is that of the Bernoulli distribution, where μ_{it} is bounded, $0 < \mu_{it} < 1$, so that the link function must map the space $\mathbb{R}$ to the interval $(0, 1)$. The literature provides a variety of link functions and the most relevant are the logit, defined by $\eta_{it} = \log\{\mu_{it}/(1-\mu_{it})\}$, and the Probit which takes the form $\eta_{it} = \Phi^{-1}(\mu_{it})$, where Φ^{-1} denotes the inverse standard Gaussian cumulative distribution function.

As mentioned above, it is crucial to specify unobserved heterogeneity in panel data models and it is also possible in the GLM framework. A common way to represent models for binary dependent variables is the latent variable formulation. The response variable y_{it} is modeled to depend on a latent continuous random variable y^*_{it}, that can be interpreted as an index of propensity for an event or an outcome to occur and consists in a linear function of the covariates, unobserved heterogeneity across individuals, and an idiosyncratic component:

$$\begin{cases} y^*_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \\ y_{it} = \mathbb{1}\{y^*_{it} \geq \tau\}, \end{cases}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function and the outcome y_{it} assumes values in 1 or 0 according to the latent variable crosses the threshold τ, which is usually fixed at 0.

Even though extremely general, models showed above follow a static specification that excludes the role played by past events. This means that the probability of an event or a choice made by an agent is assumed to be independent of the past experiences conditional on the covariates and unobserved heterogeneity. Heckman (1981a) argued about the importance of isolating the so called *true state dependence*, defined as the effect that experiencing a particular event in the present has on the probability of the same event in the future, from the *spurious state dependence*. The latter is given by the fact that, because of time-invariant unobserved heterogeneity, past experiences appear to be determinant for the probability of future events but they actually are only a proxy of the unobservable factors.

Nonlinear panel data models for binary outcomes, in both static and dynamic specifications, are of major interest in the most recent econometric literature (e.g., Hsiao, 2014; Dhaene and Jochmans, 2015; Bartolucci et al., 2016), and it has been growing fast for the last two decades. Previous reviews are provided by Arellano (2003) and Arellano and Hahn (2007). These models are also exploited in a wide range of economic applications that aim to study individual choices, such as labour market participation (Heckman and Borjas, 1980; Hyslop, 1999), portfolio choices and financial conditions of households (Alessie et al., 2004; Giarda, 2013; Brown et al., 2014), behaviour of immigrants' remittances (Bettin and Lucchetti, 2016) and firms access to credit (Pigini et al., 2016).

A wide part of literature is focusing on long panel data sets, where the larger time dimension is exploited in order to mitigate the estimation issues arising from the nonlinear formulation. In this regard, Fernández-Val and Weidner (2018) reported an extensive review concerning the estimation of the large-T panel data models. However, models for longitudinal, fixed-T, data sets are still interesting due to the large availability of panel data sets of this type. For instance, the Bank of Italy Survey on Households Income and Wealth (SHIW) is based on a rotating panel where subjects are observed three different times, while the European Union Statistics on Income and Living Conditions (EU-SILC) Database includes individuals for a period of four-years.

The aim of the present dissertation is to provide an extensive review of estimation techniques concerning nonlinear binary choice panel data models for longitudinal data. The essay will cover main theoretical aspects of the well-known random- and fixed-effects approaches. It is worth mentioning that the literature concerning the fixed-effects approach has been facing a faster growth compared with the random-effects approach. The main innovative contribution of the present work is to include the newest contributions in the literature and to compare finite sample properties of the different estimation techniques through a simulation study.

The chapter is organised in different sections: Section 1.2 introduces binary choice models and the related estimation issues. Sections 1.2.1 and 1.2.2 illustrate the main contributions of the different methodologies well known in literature of fixed- and random-effects approaches, respectively. Section 1.3 includes a simulation study comparing some estimators for the fixed-effects Dynamic Logit model. Finally, Section 1.4 concludes.

1.2 Binary choice panel data models

Binary choice panel data models deal with a response variable following a Bernoulli distribution so that y_{it} can take only two values, 1 if an event occurs or 0 otherwise. As shown in Section 1.1, it is possible to rely on the latent variable representation which can be easily adapted to binary outcomes. For ease of exposition, what follows is based on the static formulation even though most results are still valid for dynamic models. In the static case, for $i = 1, ..., n$ and $t = 1, ..., T$, the behaviour of the observable dependent variable can be described as:

$$y_{it} = \mathbb{1}\{\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} > 0\}, \tag{1.1}$$

while, for the dynamic setup, the set of explanatory variables is augmented by the lagged dependent variable and its related parameter for the state dependence, γ:

$$y_{it} = \mathbb{1}\{\alpha_i + \gamma y_{i,t-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} > 0\}, \tag{1.2}$$

and where we also assume y_{i0} to be the known initial observation for the i-th subject.

It is now straightforward to note that the expected value of y_{it}, given α_i and $\boldsymbol{x}_{it}$, equals the probability that the event occurs:

$$p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it}) = \mathrm{E}(y_{it}|\alpha_i, \boldsymbol{x}_{it}) = F(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}),$$

where $F(\cdot)$ denotes a general functional form for the inverse link function depending on the distributional assumption made on the idiosyncratic component in Equation (1.1). For instance, assuming a standard Gaussian distribution for ε_{it} we have the Probit model given by $\mathrm{E}(y_{it}|\alpha_i, \boldsymbol{x}_{it}) = \Phi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})$ where $\Phi(\cdot)$ denotes the standard Gaussian distribution function (cdf), while for a standard logistic cdf we get the Logit model where

$$\mathrm{E}(y_{it}|\alpha_i, \boldsymbol{x}_{it}) = \frac{\exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})}{1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})}.$$

Based on the above formulations, it is now possible to define the likelihood function for the sample, $\mathscr{L}(\cdot)$. Assuming a collection of independent and identically distributed observations from a Bernoulli random variable we define the probability function for an observation y_{it} as

$$p(y_{it}|\alpha_i, \boldsymbol{x}_{it}) = F(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})^{y_{it}}[1 - F(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]^{1-y_{it}}, \tag{1.3}$$

and the likelihood function based on Equation (1.3) as:

$$\mathscr{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{t=1}^{T} p(y_{it}|\alpha_i, \boldsymbol{x}_{it}). \tag{1.4}$$

Additional assumptions on the individual effects are needed in order to obtain consistent estimators for $\boldsymbol{\beta}$ since the likelihood function depends on the unobserved heterogeneity, which is represented by α_i and it is common to all observations for the i-th individual that are no longer i.i.d. by construction.

The simplest approach is to assume absence of unobserved heterogeneity, given by the condition $\alpha_i = \alpha$ for $i = 1, ..., n$, so that α enters the likelihood function, $\mathscr{L}(\alpha, \boldsymbol{\beta})$, as a parameter to be estimated along with $\boldsymbol{\beta}$. Under this setup a consistent estimator is obtained via the classical maximisation of the log-likelihood function, $\ell(\cdot) = \log \mathscr{L}(\cdot)$, given by the first order conditions of null score function denoted as

$$\begin{aligned}\frac{\partial \ell(\alpha, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \boldsymbol{s}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \alpha) = \mathbf{0},\\ \frac{\partial \ell(\alpha, \boldsymbol{\beta})}{\partial \alpha} &= s_{\alpha}(\boldsymbol{\beta}, \alpha) = 0.\end{aligned}$$

The resulting estimators will be denoted as $\hat{\alpha}_{ML}$ and $\hat{\boldsymbol{\beta}}_{ML}$. Although this procedure is standard and simple to implement, it is based on the assumption of no unobserved heterogeneity which is often unrealistic in empirical applications. Moreover, the estimator is not consistent for parameters $\boldsymbol{\beta}$ when the homogeneity hypothesis is violated (Wooldridge, 2010).

Literature provides two main branches. Section 1.2.1 presents the variety of techniques concerning the fixed-effects approach, where unobserved heterogeneity is expressed by a set of unrelated fixed parameters across individuals. Then, the random-effects approach is illustrated in Section 1.2.2 and it consists in assuming the individual effects to follow a random variable.

1.2.1 Fixed-effects approach

This approach consists in assuming that the individual specific effects, α_i, are fixed parameters. The main advantage with respect to the random-effects approach is that there is no need for distributional assumptions and individual effects can be freely correlated with the covariates. Under this setup, the α_i-s enter the likelihood function in Equation (1.4) as parameters to be estimated. Due to this setup, we can only include time-varying explanatory variables in order to achieve identification. Moreover, the standard maximum likelihood estimation of the whole set of parameters leads to inconsistent estimators because of incidental parameters problem (Neyman and Scott, 1948).

Two different literature branches have developed in order to overcome this problem, target-adjusted estimators and conditional inference. The aim of the first approach is to reduce the leading bias component of the maximum likelihood estimator via different techniques. Prominent contributions are provided by Hahn and Newey (2004), Carro (2007), Fernández-Val (2009), Dhaene and Jochmans (2015) and Bartolucci et al. (2016). The second approach considers conditional probabilities exploiting sufficient statistics for the incidental parameters. Main contributions in this field are by Chamberlain (1980) for the static version of the model and by Chamberlain (1993), Honoré and Kyriazidou (2000) and Bartolucci and Nigro (2010, 2012) for the dynamic setup.

The incidental parameters problem

As previously mentioned, in order to take unobserved heterogeneity into account, it could be reasonable to enlarge the set of parameters to be estimated in $(\alpha_1, ..., \alpha_n, \boldsymbol{\beta}')'$ including in the set of the regressors a dummy for each subject in the sample. As a result, this situation leads to the well known *incidental parameters problem* described by Neyman and Scott (1948). This problem can be seen as the asymmetry in information provided by data. As a matter of fact, only additional observations over time for an individual i give information on α_i, while new individuals in the data set increase the number of parameters to be estimated. Hence, the maximum likelihood estimator for individual intercepts, $\hat{\alpha}_{iML}$, requires $T \to \infty$ in order to be consistent for α_i. Moreover, the inconsistent estimation of individual effects affects other parameters estimates.

Lancaster (2000) provides an intuitive illustration. Consider a random variable y_{it} following a Gaussian distribution such that $y_{it} \sim N(\alpha_i, \sigma_0^2)$. The ML estimator for the mean parameters is given by $\hat{\alpha}_i = \frac{1}{T}\sum_t y_{it}$, while for the variance parameter by $\hat{\sigma^2} = \frac{1}{nT}\sum_i \sum_t (y_{it} - \hat{\alpha}_i)^2$ which is proven to be inconsistent for $n \to \infty$ and fixed T since it converges to $\frac{T-1}{T}\sigma_0^2$. Thus, the MLE for the parameter σ_0^2 is not consistent due to the limited set of observation over T for each $\hat{\alpha}_i$.

In order to understand the bias of the ML estimator due to the incidental parameters problem for binary choice models, it is useful to recall the description provided in Arellano and Hahn (2007). The starting point is that, when T is fixed, the ML estimator of common parameters, $\hat{\boldsymbol{\beta}}$, shows a bias of order $1/T$, denoted $\mathbb{B}/T$ for some $\mathbb{B}$, so that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \frac{\mathbb{B}}{T} + O(1/T^2),$$

where $\boldsymbol{\beta}_0$ is the value of the estimator when T tends to infinity and $O(1/T^2)$ denotes higher-order bias components.

The bias component is also present in the asymptotic distribution of the estimator. The estimator $\hat{\boldsymbol{\beta}}$ converges to his probability limit $\boldsymbol{\beta}_*$, which is different from the true value of parameters. It is also possible to notice the asymptotic bias of $\hat{\boldsymbol{\beta}}$ considering the asymptotic distribution of $\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, which is not centered on $\mathbf{0}$. Given the condition of n and T growing at the same rate ρ, $n/T \to \rho$, we obtain

$$\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*) + \sqrt{nT}\frac{\mathbb{B}}{T} + O(\sqrt{n/T^3}) \xrightarrow{d} N(\mathbb{B}\sqrt{\rho}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega}$ denotes the variance-covariance matrix.

The bias also arises in the expected score function of the profile likelihood, where the nuisance parameters, α_i, are concentrated out. This effect can be formalised considering the first order condition

$$\frac{1}{nT}\sum_{i=1}^{n}\boldsymbol{s}_{\beta,i}(\hat{\alpha}_i(\boldsymbol{\beta}),\boldsymbol{\beta})=\mathbf{0}. \tag{1.5}$$

The bias of the ML estimator arises because these conditions cannot be satisfied. When evaluated at the true value of parameters $\boldsymbol{\beta}=\boldsymbol{\beta}_0$, with $N\rightarrow\infty$ and T fixed, the l.h.s. of Equation (1.5) does not converge in probability to $\mathbf{0}$ because $\hat{\alpha}_i(\boldsymbol{\beta}_0)$ does not converge to its true value α_{i0}. Therefore, we define a bias component $b_i(\boldsymbol{\beta}_0)$ of order $(1/T)$ and a residual component with order $o(1/T)$ in the expected score

$$\mathrm{E}\left[\boldsymbol{s}_{\beta,i}(\boldsymbol{\beta}_0,\hat{\alpha}_i(\boldsymbol{\beta}_0))/T\right]=\frac{b_i(\boldsymbol{\beta}_0)}{T}+o\left(\frac{1}{T}\right).$$

Finally, it is possible to consider the bias of the profile likelihood. Again, since the MLE for α_i does not converge to its true value, maximising $\ell_i(\boldsymbol{\beta})=\ell_i(\boldsymbol{\beta},\hat{\alpha}_i(\boldsymbol{\beta}))$ yields inconsistent estimates of parameters. Consider now the infeasible profile likelihood function given by

$$\bar{\ell}_i(\boldsymbol{\beta})=\ell_i(\boldsymbol{\beta},\bar{\alpha}_i(\boldsymbol{\beta})), \tag{1.6}$$

where $\bar{\alpha}_i(\boldsymbol{\beta})$ is the ML for α_i when $T\rightarrow\infty$, such that $\bar{\alpha}_i(\boldsymbol{\beta}_0)=\alpha_{i0}$. The profile likelihood in Equation (1.6) can be considered a target function since its maximand results to be an unbiased estimator for the parameters of interest. Hence, the bias of the profile likelihood, $B_i(\boldsymbol{\beta})$, arises from the expectation of the difference

$$\mathrm{E}\left[\ell_i(\boldsymbol{\beta},\hat{\alpha}_i(\boldsymbol{\beta}))/T-\ell_i(\boldsymbol{\beta},\bar{\alpha}_i(\boldsymbol{\beta}))/T\right]=\frac{B_i(\boldsymbol{\beta})}{T}+o\left(\frac{1}{T}\right).$$

Target-corrected estimators

This is a class of estimators that deals with the asymptotic bias of the ML estimator. Arellano and Hahn (2007) classified these techniques in three main categories: bias-corrected estimators, correction of the moment equation, and corrected objective-function estimators. In general, the underlying idea of this approach is to mitigate the bias of MLE, lowering its order from $(1/T)$ to $(1/T^2)$. The advantage of this approach is given by its applicability. Indeed, the results are extremely general and are easily adaptable to binary choice models, both static and dynamic and regardless the functional form assumed for the error term. On the contrary, a possible drawback is given by the fact that the ML estimator is used to compute the bias component and it may influence the goodness of estimation when T is small.

Extensive reviews of these techniques are provided by Arellano (2003) and Arellano and Hahn (2007). We now consider the more recent literature, where several methods have been proposed to bypass the incidental parameters problem. We can consider, following Arellano and Hahn (2007), three main groups: the first type has to do directly with asymptotic bias of the estimator and the main contributions are given by Hahn and Newey (2004), Fernández-Val (2009), Hahn and Kuersteiner (2011), and Dhaene and Jochmans (2015). A second approach is given by the correction of the first order conditions of a modified likelihood function as proposed by Carro (2007). Finally, the third group deals with target functions, generally modifications of the (profile) likelihood function, as argued by Bester and Hansen (2009), Arellano and Hahn (2016), and Bartolucci et al. (2016).

Since the following proposals exploit a wide range of techniques, we need a small preamble in order to recall and clarify the notation. First of all, we define the parameters. As we will see, some estimators deals with static models, some with dynamic, others with both. Therefore, following the notation in Section 1.2, we define α_i as the individual heterogeneity or nuisance parameters. Furthermore, $\boldsymbol{\beta}$ will be the set of parameters related to the regressors in static models, γ the state dependence parameter in dynamic models. It is useful to define $\boldsymbol{\theta}' = \{\boldsymbol{\beta}', \gamma\}$, a vector collecting all the parameters of interest in the dynamic setup.

Secondly, we define the log-likelihood function as

$$\ell(\alpha_1, ..., \alpha_n, \boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{t=1}^{T} \ell_{it}(\alpha_i, \boldsymbol{\theta})$$

and the individual contributions as $\ell_i(\alpha_i, \boldsymbol{\theta}) = \sum_{t=1}^{T} \ell_{it}(\alpha_i, \boldsymbol{\theta})$. The profile likelihood function is given by $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta})$ where nuisance parameters $(\alpha_1, ..., \alpha_n)$ are concentrated out, or more precisely, evaluated at its MLE for a given value of $\boldsymbol{\theta}$. In general, we denote the ML estimator of parameter vector $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}$. Unless differently specified, we take into consideration the maximisation of the profile likelihood to obtain $\hat{\boldsymbol{\theta}}$. Finally, we try to keep the notation homogeneous throughout the dissertation, but at the same time as close as possible to the original one of the different contributions.

We first examine the bias-corrected estimators. A seminal paper in this branch of literature is the one of Hahn and Newey (2004). The authors propose two different ways to obtain a measure for the distortion of the MLE for static nonlinear panel models: they derive an analytical expression for the bias and a jackknife estimator. The former is computed via sample counterparts of the moments employed in the bias formula,[1] denoted by $\hat{\mathbb{B}}$. The idea is to correct the asymptotic limit of the ML estimator subtracting the estimated bias, so that

$$\hat{\boldsymbol{\beta}}^1 = \hat{\boldsymbol{\beta}} - \hat{\mathbb{B}}(\hat{\boldsymbol{\beta}}), \tag{1.7}$$

where $\hat{\boldsymbol{\beta}}^1$ is the adjusted estimator. However, since $\hat{\boldsymbol{\beta}}$ is used in order to compute $\hat{\mathbb{B}}$, when T is small the bias of the MLE could severely spread to the estimator $\hat{\mathbb{B}}$. In this situation authors proposed to iterate the estimation procedure in Equation (1.7) updating the estimation of $\mathbb{B}$ such that $\hat{\boldsymbol{\beta}}^k = \hat{\boldsymbol{\beta}} - \hat{\mathbb{B}}(\hat{\boldsymbol{\beta}}^{k-1})$. Moreover, the resulting estimator $\hat{\boldsymbol{\beta}}^\infty$ exhibits higher finite sample performances. Hahn and Kuersteiner (2011) extend the results of Hahn and Newey (2004) to general dynamic nonlinear panel models deriv-

[1]See Hahn and Newey (2004) for further details.

ing regularity conditions for the estimator. Due to the dynamic setup, the covariates are assumed to be stationary and independent across the subjects in the sample. Another contribution in this field is given by Fernández-Val (2009), who derives the analytical bias formula for binary choice model, both static and dynamic, with predetermined regressors and proposed an equivalent estimator based on the adjustment of the score function.

Jackknife techniques also play an important role among the bias-corrected estimators. The proposal of Hahn and Newey (2004) is extremely simple and it is given by

$$\tilde{\boldsymbol{\beta}} = T\hat{\boldsymbol{\beta}} - (T-1)\sum_{t=1}^{T}\hat{\boldsymbol{\beta}}(t)/T,$$

where $\hat{\boldsymbol{\beta}}(t)$ is the ML estimator computed subtracting the t-th observation from the sample. The authors prove that the order of the bias of $\tilde{\boldsymbol{\beta}}$ is lowered to $1/T^2$. Dhaene and Jochmans (2015) provide two different estimators based on the split-panel jackknife for dynamic nonlinear models. A peculiarity of this methodology is given by the fact that it allows for generated regressors, which typically arise in case of endogeneity and sample selection. This approach imposes some regularity conditions on the data, namely, stationarity, a sufficient degree of mixing, and independence of observations across subjects. The procedure is based on the idea of considering sub-panels, consisting of a reduced number of consecutive observations over time for each subject in order to preserve the dynamic structure of the data. This idea can be formalised as follows. Consider a subset of consecutive observations $S \subsetneq \{1, ..., T\}$ such that $|S| \leq T_{min}$, where $|S|$ denotes the cardinality of the subset S and T_{min} is the least number of observations for which the ML estimator exists. Define $\boldsymbol{\theta} = (\boldsymbol{\beta}, \gamma)$ as the vector of parameters to be estimated maximizing the profile likelihood, $\ell(\boldsymbol{\theta})$, where nuisance parameters $(\alpha_1, ..., \alpha_n)$ are concentrated out. Authors showed that it is possible to consistently estimate the leading bias component of the MLE as following

$$\frac{|S|}{T-|S|}(\hat{\boldsymbol{\theta}}_S - \hat{\boldsymbol{\theta}}),$$

where $\hat{\boldsymbol{\theta}}_S$ is the ML estimator of $\boldsymbol{\theta}$ based on the subset of observations considered in S and $\hat{\boldsymbol{\theta}}$ is the ML estimator considered on the full data set. This kind of estimator could suffer from the arbitrary choice of the sub-panel S. Define an integer number $g \geq 2$ such that $T \geq g \cdot T_{min}$, where T_{min} is the minimum number of observations for which the ML estimator exists, and suppose to split the panel into a collection of g non-overlapping sub-panels $\mathbb{S} = \{S_1, ..., S_g\}$. A consistent estimator for the bias can be obtained by averaging over the subsets S in $\mathbb{S}$ the estimator $\hat{\boldsymbol{\theta}}_S$ so that

$$\frac{1}{g-1}(\bar{\boldsymbol{\theta}}_{\mathbb{S}} - \hat{\boldsymbol{\theta}}),$$

where

$$\bar{\boldsymbol{\theta}}_{\mathbb{S}} = \sum_{S \in \mathbb{S}} \frac{|S|}{T} \hat{\boldsymbol{\theta}}_S.$$

The bias-corrected estimator, $\tilde{\boldsymbol{\theta}}$, is derived subtracting from the MLE the estimated bias, as follows

$$\tilde{\boldsymbol{\theta}} = \frac{g}{g-1}\hat{\boldsymbol{\theta}} - \frac{1}{g-1}\bar{\boldsymbol{\theta}}_{\mathbb{S}} \tag{1.8}$$

As an example, consider the situation where $g = 2$, $T_{min} = 3$ so that, in order to satisfy the feasibility condition of $T \geq 6$, set $T = 10$. This setup consists in taking into account the two half-panels given by

$$\mathbb{S} = \{S_1, S_2\} \qquad \text{where} \qquad S_1 = \{1, 2, 3, 4, 5\}, S_2 = \{6, 7, 8, 9, 10\}.$$

First of all, we have to compute $\hat{\boldsymbol{\theta}}_{S_1}$, $\hat{\boldsymbol{\theta}}_{S_2}$ and $\hat{\boldsymbol{\theta}}$, maximizing the profile log-likelihood computed in the three different time spans. The next step consists in computing the average over the element of $\mathbb{S}$ given by $\bar{\boldsymbol{\theta}}_{\mathbb{S}} = \frac{1}{2}(\hat{\boldsymbol{\theta}}_{S_1} + \hat{\boldsymbol{\theta}}_{S_2})$. Finally, the bias-corrected estimator is computed as

$$\tilde{\boldsymbol{\theta}} = 2\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbb{S}}.$$

However, it could happen that splitting the panel in g sub-samples leads to partitions of different cardinality. In order to avoid discretion in the allocation of the observations over the different sub-classes, define an equivalence class $\{\mathbb{S}_1, ..., \mathbb{S}_m\}$ sharing the same set of cardinalities of elements in $\mathbb{S}$. Dhaene and Jochmans propose to average the estimator $\bar{\boldsymbol{\theta}}_{\mathbb{S}}$ over the equivalence class of $\mathbb{S}$,

$$\bar{\boldsymbol{\theta}} = \frac{1}{m}\sum_{j=1}^{m} \bar{\boldsymbol{\theta}}_{\mathbb{S}_j},$$

replacing $\bar{\boldsymbol{\theta}}_{\mathbb{S}}$ in Equation (1.8) with $\bar{\boldsymbol{\theta}}$.

Consider now the previous example in the situation of $g = 3$. In this case $\mathbb{S} = \{S_1, S_2, S_3\}$ and its equivalence class is given by

$$\begin{aligned}
\mathbb{S}_1 &= \{S_{11}, S_{12}, S_{13}\}, \quad \text{where} \quad S_{11} = \{1,2,3,4\}, S_{12} = \{5,6,7\}, S_{13} = \{8,9,10\}\\
\mathbb{S}_2 &= \{S_{21}, S_{22}, S_{23}\}, \quad \text{where} \quad S_{11} = \{1,2,3\}, S_{12} = \{4,5,6,7\}, S_{13} = \{8,9,10\}\\
\mathbb{S}_3 &= \{S_{31}, S_{32}, S_{33}\}, \quad \text{where} \quad S_{11} = \{1,2,3\}, S_{12} = \{4,5,6\}, S_{13} = \{7,8,9,10\},
\end{aligned}$$

so that it is now required to compute three bias estimators, namely, $\bar{\boldsymbol{\theta}}_{\mathbb{S}_1}$, $\bar{\boldsymbol{\theta}}_{\mathbb{S}_2}$ and $\bar{\boldsymbol{\theta}}_{\mathbb{S}_3}$. We have now to plug the average given by $\bar{\boldsymbol{\theta}} = \frac{1}{3}(\bar{\boldsymbol{\theta}}_{\mathbb{S}_1} + \bar{\boldsymbol{\theta}}_{\mathbb{S}_2} + \bar{\boldsymbol{\theta}}_{\mathbb{S}_3})$ in the estimator in Equation (1.8), getting

$$\tilde{\boldsymbol{\theta}} = \frac{3}{2}\hat{\boldsymbol{\theta}} - \frac{1}{2}\bar{\boldsymbol{\theta}}.$$

Dhaene and Jochmans (2015) also prove that the aforementioned split-panel jackknife procedure can be exploited in order to correct the profile likelihood function instead of the estimator. A consistent bias-adjusted estimator is then obtained maximising the resulting modified function. Moreover, an interesting extension of the analytical corrections and the split-panel jackknife (Dhaene and Jochmans, 2015) for individual and time effects is provided by Fernández-Val and Weidner (2016).

A second way in order to mitigate the bias of the MLE is to consider the estimating equation as proposed by Carro (2007). The incidental parameters problem affects the first order conditions of the concentrated log-likelihood, which are not centered at zero. In order to recenter the score function Carro considered a modified version of the likelihood function. Cox and Reid (1987) propose a reparametrised version of the of the likelihood in order to obtain orthogonality between nuisance parameters and parameters of interest. Arellano (2003) shows how to exploit the modified likelihood for static binary choice models and that the modified profile likelihood can be rewritten in terms of the original parameters. In this framework, Carro (2007) proposes a modified estimating equation for dynamic models, given by

$$\begin{aligned} M\boldsymbol{s}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}) =& \boldsymbol{s}_{\alpha_i}(\boldsymbol{\theta},\hat{\alpha}_i(\boldsymbol{\theta})) - \frac{1}{2}\frac{1}{s_{\alpha\alpha_i}(\boldsymbol{\theta},\hat{\alpha}_i(\boldsymbol{\theta}))}\left(\boldsymbol{s}_{\boldsymbol{\theta}\alpha\alpha_i}(\hat{\alpha}_i(\boldsymbol{\theta})) + s_{\alpha\alpha\alpha_i}(\hat{\alpha}_i(\boldsymbol{\theta}))\frac{\partial\hat{\alpha}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) \\ &+ \frac{\partial}{\partial\alpha_i}\left(\frac{1}{\mathrm{E}(s_{\alpha\alpha_i}(\boldsymbol{\theta},\alpha_i))}\mathrm{E}(\boldsymbol{s}_{\boldsymbol{\theta}\alpha_i}(\boldsymbol{\theta},\alpha_i))\right)\Bigg|_{\alpha_i=\hat{\alpha}_i(\boldsymbol{\theta})} = \mathbf{0}, \qquad (1.9) \end{aligned}$$

where the subscripts in the score function $s(\cdot)$ denotes the derivatives of the profile likelihood w.r.t. a parameter, e.g., $s_{\alpha\alpha_i} = \partial^2\ell(\boldsymbol{\theta},\alpha_i(\boldsymbol{\theta}))/\partial\alpha_i^2$. The estimator $\hat{\boldsymbol{\theta}}_{MMLE}$ that satisfies Equation (1.9) shares the same asymptotic properties of the MLE except for the reduced order of the bias.

The third class of bias-corrected estimators deals with objective functions. Bester and Hansen (2009) developed a penalty function for the unconstrained likelihood function, differently from other approaches who exploits the profile likelihood. Define now the penalised objective function by

$$Q(\alpha_1,...,\alpha_n,\boldsymbol{\theta}) = \sum_{i=1}^{n}\ell_i(\alpha_i,\boldsymbol{\theta}) - \pi_i(\alpha_i,\boldsymbol{\theta}),$$

whose maximand results being the bias-adjusted estimator. A crucial role is played by the penalty function $\pi_i(\alpha_i,\boldsymbol{\theta})$, defined as

$$\pi_i(\alpha_i,\boldsymbol{\theta}) = \frac{1}{2}\mathrm{trace}(-\hat{\boldsymbol{I}}_{\alpha_i}^{-1}\hat{\boldsymbol{V}}_{\alpha_i}) - \frac{k}{2},$$

where $k = \dim(\alpha_i)$ is a parameter,[2] the elements $\hat{\boldsymbol{I}}_{\alpha_i}$ and $\hat{\boldsymbol{V}}_{\alpha_i}$ are respectively the information matrix for the parameter α_i and a heteroskedasticity and autocorrelation robust estimator for the variance of the expected score,

[2]This approach can be easily extended to multiple effects, so that $k \geq 1$.

formally we have

$$\hat{\boldsymbol{I}}_{\alpha_i} = s_{\alpha\alpha_i}(\alpha_i, \boldsymbol{\theta})/T,$$
$$\hat{\boldsymbol{V}}_{\alpha_i} = \sum_{t=-m}^{m} \sum_{t=\max(1,t)}^{\max(T,T+l)} s_{\alpha_i,t}(\alpha_i, \boldsymbol{\theta}) s_{\alpha_i,t-l}(\alpha_i, \boldsymbol{\theta})',$$

where m is a bandwidth parameter, and the additional subscript in the score function indicates the i,t-th observation-specific contribution to the individual score. The main advantages of this approach are its wide applicability for static and dynamic models and the computational easiness, since it requires only the computation of the score function and the Hessian matrix. Authors argue about the asymptotic equivalence of their approach with those previously discussed, highlighting the trade-off between the generality of their proposal and the better finite sample properties of other model-specific estimators like Fernández-Val (2009) and Carro (2007). Similarly, Arellano and Hahn (2016) propose two corrections for the profile likelihood. The so called "trace-based" correction is not restricted to the likelihood setting and is extremely close to the methodology proposed by Bester and Hansen (2009). Their second proposal is the "determinant-based" correction that exploits the log determinants of $\hat{\boldsymbol{I}}_{\alpha_i}$ and $\hat{V}_{\alpha_i}$.

Furthermore, in this field, Bartolucci et al. (2016) propose a modified profile likelihood as objective function. In general, the modified profile likelihood, $\ell_{M,i}(\cdot)$, takes the form

$$\ell_{M,i}(\hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta}) = \ell_i(\hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta}) + M_i(\hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta})$$

where $M_i(\cdot)$ denotes the adjustment function. The authors exploit the modification proposed by Severini (1998) given by

$$M_i(\boldsymbol{\theta}) = \frac{1}{2} \log|-s_{\alpha_i\alpha_i}(\hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta})| - \log|\boldsymbol{I}_{\alpha_i\alpha_i}(\hat{\alpha}_i, \hat{\boldsymbol{\theta}}; \hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta}|,$$

where $s_{\alpha_i\alpha_i}$ denotes the second derivative of the concentrated log-likelihood w.r.t. the parameter α_i and

$$\boldsymbol{I}_{\alpha_i\alpha_i}(\hat{\alpha}_i, \hat{\boldsymbol{\theta}}; \hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta}) = \mathrm{E}_{\{\hat{\alpha}_i, \hat{\boldsymbol{\theta}}\}}[s_{\alpha_i}(\hat{\alpha}_i, \hat{\boldsymbol{\theta}}) s_{\alpha_i}(\hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta})]$$

is the expected value of the product of the score w.r.t. α_i where the first element is evaluated at the ML estimates.

The application to dynamic nonlinear panel binary choice model is of main interest. Consider now the model in Equation (1.3) in its dynamic setup. The first term of the adjustment can be derived analytically, such that

$$-s_{\alpha_i\alpha_i}(\boldsymbol{\theta}, \hat{\alpha}_i(\boldsymbol{\theta})) = \sum_t \left[\frac{f(\tilde{\mu}_{it})^2}{F(\tilde{\mu}_{it})[1 - F(\tilde{\mu}_{it})]} - C(\tilde{\mu}_{it}) \right],$$

where

$$C(\tilde{\mu}_{it}) = (y_{it} - F(\tilde{\mu}_{it})) \left[\frac{f(\tilde{\mu}_{it})}{F(\tilde{\mu}_{it})[1 - F(\tilde{\mu}_{it})]} - \frac{f(\tilde{\mu}_{it})^2(1 - 2F(\tilde{\mu}_{it}))}{F(\tilde{\mu}_{it})^2[1 - F(\tilde{\mu}_{it})]^2} \right].$$

In this formulation, $f(\cdot)$ denotes the density derived from the distribution function $F(\cdot)$ and $\tilde{\mu}_{it} = \hat{\alpha}_i(\boldsymbol{\theta}) + \gamma y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta}$ is the linear predictor we obtain considering $\alpha_i = \hat{\alpha}_i(\boldsymbol{\theta})$.

Unfortunately, the term $\boldsymbol{I}_{\alpha_i\alpha_i}(\hat{\alpha}_i, \hat{\boldsymbol{\theta}}; \hat{\alpha}_i(\boldsymbol{\theta}), \boldsymbol{\theta})$ does not admit a closed-form expression. Authors proposed two different ways in order to obtain this element. The first exploits all the possible configurations of the vector $(y_{i1}, ..., y_{iT})$, weighting the product of the scores over the probability assigned to each vector configuration. However, this technique is convenient when the time dimension of the data set is moderate. The second proposal consists in using a Monte Carlo approximation that makes use of simulations from the model and has a moderate computational cost.

Conditional inference

The conditional inference approach exploits the existence of sufficient statistics for the incidental parameters. The key point of these techniques is the possibility to make inference independently from the nuisance parameters. However, as mentioned above, sufficient statistics are always not available and solutions are model specific. For binary choice models, consider the joint probability for $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT})$, which is given by the product of the probabilities of each observation over time of the same subject. The probability is denoted as $p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i)$ where, again, $\boldsymbol{X}_i$ is the matrix collecting the related set of covariates $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})$. Consider now a statistic h_i with probability distribution $p(h_i|\alpha_i, \boldsymbol{X}_i)$. If conditioning the joint probability for $\boldsymbol{y}_i$ on h_i leads to a distribution that is independent of α_i, then h_i is said to be a sufficient statistic for the incidental parameters:

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i,\ h_i) = \frac{p(\boldsymbol{y_i}|\alpha_i, \boldsymbol{X}_i)}{p(h_i|\alpha_i, \boldsymbol{X}_i)}. \tag{1.10}$$

Andersen (1970) shows that the maximand of the log-likelihood function based on the conditional density in Equation (1.10) is a consistent estimator for parameters of interest. Although this idea looks simple and intuitive, it may happen that a sufficient statistic does not exist or it is not trivial to identify for general binary choice models (Hsiao, 2014).

A specification admitting a sufficient statistic is the Logit model (Hsiao, 2014; Cameron and Trivedi, 2005). The probability function for the Logit model can be written as

$$p(y_{it}|\alpha_i, \boldsymbol{x}_{it}) = \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]}{1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})},$$

and the joint probability for $\boldsymbol{y}_i$ is

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i) = \prod_{t=1}^{T} p(y_{it}|\alpha_i, \boldsymbol{x}_{it}) = \frac{\exp[\alpha_i y_{i+} + (\sum_{t=1}^{T} y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\prod_{t=1}^{T}[1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]}, \tag{1.11}$$

where the total score $y_{i+} = \sum_{t=1}^{T} y_{it}$ is the sufficient statistic for the incidental parameter α_i (Andersen, 1970). This can be proven in a simple way, following Cameron and Trivedi (2005). The probability in Equation (1.11) must be conditioned on the sufficient statistic y_{i+}. From Bayes' rule, the conditional probability of the configuration $\boldsymbol{y}_i$ given α_i, $\boldsymbol{X}_i$ and y_{i+} can be expressed as

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i+}) = \frac{p(y_{i+}|\alpha_i, \boldsymbol{X}_i, \boldsymbol{y}_i)p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i)}{p(y_{i+}|\alpha_i, \boldsymbol{X}_i)}.$$

Since the total score y_{i+} is the sum of the elements in $\boldsymbol{y}_i$, the probability $p(y_{i+}|\alpha_i, \boldsymbol{X}_i, \boldsymbol{y}_i) = 1$ by definition. Therefore, it is possible to write

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i+}) = \frac{p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i)}{p(y_{i+}|\alpha_i, \boldsymbol{X}_i)}. \tag{1.12}$$

The numerator in Equation (1.12) is given by Equation (1.11), but we have to define $p(y_{i+}|\alpha_i, \boldsymbol{X}_i)$. This probability is given by the sum of the probabilities of observing each possible vector configuration of binary responses $\boldsymbol{z} = (z_1, ..., z_T)$ such that $z_+ = y_{i+}$, where $z_+ = \sum_{t=1}^T z_t$, which is

$$p(y_{i+}|\alpha_i, \boldsymbol{X}_i) = \frac{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp(\alpha_i z_+) \exp[(\sum\limits_{t=1}^T z_t \boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\prod_{t=1}^T [1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]}.$$

Finally, it is possible to compute the conditional probability of $p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+})$, independent of the parameter α_i as follows

$$\begin{aligned} & p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i+}) \\ = & \frac{\exp(\alpha_i y_{i+}) \exp[(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\prod_{t=1}^T [1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]} \frac{\prod_{t=1}^T [1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp(\alpha_i z_+) \exp\left[\left(\sum_{t=1}^T z_t \boldsymbol{x}_{it}\right)' \boldsymbol{\beta}\right]}, \\ = & \frac{\exp[(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp\left[\left(\sum_{t=1}^T z_t \boldsymbol{x}_{it}\right)' \boldsymbol{\beta}\right]} = p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}). \end{aligned} \tag{1.13}$$

Equation (1.13) defines the Conditional Logit model shown in McFadden (1974) and Chamberlain (1980). Given this result, it is now possible to set the conditional log-likelihood function, as the sum of the logarithm of the individual probabilities

$$\ell(\boldsymbol{\beta}) = \sum_i \mathbb{1}\{0 < y_{i+} < T\} \log p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}), \tag{1.14}$$

excluding individuals characterized by a total score of 0 or T, because their conditional log-probability is equal to 0 by construction. The function in Equation (1.14) can be maximised with respect to $\boldsymbol{\beta}$ by the Newton-Raphson algorithm, obtaining the Conditional Maximum Likelihood (CML) estimator $\hat{\boldsymbol{\beta}}_{CML}$.

Differently from the static case, conditional inference for dynamic models is more difficult because a sufficient statistic is not always available. Different approaches have been implemented in order to overcome this problem. Consider the model given in Equation (1.2) where ε_{it} is logistically distributed, defining the Dynamic Logit model (DL) (Hsiao, 2014), where the probability for a response is

$$p(y_{it}|\alpha_i, \boldsymbol{x}_{it}, y_{i0}, ..., y_{i,t-1}) = \frac{\exp[y_{it}(\gamma y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i)]}{1 + \exp(\gamma y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i)}, \tag{1.15}$$

and y_{i0} is the initial observation assumed to be known. In this case, the probability for the response configuration $(y_{i1}, ..., y_{iT})$ is

$$p(\boldsymbol{y}_i|\alpha_i, y_{i0}, \boldsymbol{X}_i) = \frac{\exp[y_{i+}\alpha_i + \left(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it}\right)' \boldsymbol{\beta} + y_{i*}\gamma]}{\prod_{t=1}^T [1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \gamma y_{i,t-1})]}, \tag{1.16}$$

where $y_{i*} = \sum_{t=1}^T y_{i,t-1}y_{i,t}$. It can be proven that the total score y_{i+} is no longer a sufficient statistic for the incidental parameters as in the static specification.

A first feasible solution for the DL model is given by Chamberlain (1993), even though quite restrictive. Consider the DL given in Equation (1.15) where the exogenous variables, $\boldsymbol{x}_{it}$, are excluded and $T \geq 3$. It is possible to make inference on the parameter γ independently of incidental parameters via a conditional approach. Given these assumptions, as an example, when

$T = 3$ it is possible to set the probabilities for a vector $(y_{i0}, ..., y_{i3})$ as follows:

$$\begin{aligned} p(y_{i0} = 1|\alpha_i) &= P_0(\alpha_i), \\ p(y_{it} = 1|\alpha_i, y_{i0}, ..., y_{i,t-1}) &= \frac{\exp(\gamma y_{i,t-1} + \alpha_i)}{1 + \exp(\gamma y_{i,t-1} + \alpha_i)} \qquad t = 1, ..., T., \end{aligned}$$

Under this setup, the probability for $(y_{i0}, ..., y_{i3})$ is independent of α_i conditional on $y_{i1} + y_{i2} = 1$. Maximising the resulting conditional log-likelihood function

$$\ell = \sum_{i=1}^{n} \mathbb{1}(y_{i1} + y_{i2} = 1)\{y_{i1}[\gamma(y_{i0} - y_{i3})] - \log[1 + \exp(y_{i0} - y_{i3})]\}$$

yields a $\sqrt{n}$ consistent estimator of γ.

The approach of Chamberlain (1993) has been extended by Honoré and Kyriazidou (2000). The authors allow for a specification that includes a set of strictly exogenous variables as regressors, in addition to the unobserved heterogeneity and the lag of the dependent variable, following a DL model. Moreover, they showed how to identify and estimate $\boldsymbol{\beta}$ and γ independently of α_i. Given $T = 3$, this can be done by maximizing a weighted conditional log-likelihood function given by

$$\sum_{i=1}^{n} \mathbb{1}(y_{i1} + y_{i2} = 1) K\left(\frac{\boldsymbol{x}_{i2} - \boldsymbol{x}_{i3}}{\sigma_n}\right) \log p(\boldsymbol{y_i}|\alpha_i, \boldsymbol{X}_i, y_{i0}, y_{i1} + y_{i2} = 1, y_{i3}, \boldsymbol{x}_{i2} = \boldsymbol{x}_{i3}),$$

where $K(\cdot)$ is a kernel density function, carefully chosen, used in order to weigh observations. In particular, weights are inversely proportional to the magnitude of the difference $(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i3})$, σ_n is a fixed bandwidth that depends on n and

$$\begin{aligned} & p(\boldsymbol{y_i}|\alpha_i, \boldsymbol{X}_i, y_{i0}, y_{i1} + y_{i2} = 1, y_{i3}, \boldsymbol{x}_{i2} = \boldsymbol{x}_{i3}) \\ &= \frac{\exp\{y_{i1}[(\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})'\boldsymbol{\beta} + \gamma(y_{i0} - y_{i3})]\}}{1 + \exp[(\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})'\boldsymbol{\beta} + \gamma(y_{i0} - y_{i3})]}, \end{aligned}$$

where y_{i0} and y_{i3} can be either 0 or 1. Despite the proposed estimator is proven to be consistent and asymptotically normal, it shows some drawbacks.

The convergence rate, due to the presence of the kernel density function, is slower than $\sqrt{n}$ and the conditions exploited for identification, namely that $y_{i1} + y_{i2} = 1$, and the weight given by the kernel, limit the number of individuals that actually contribute to the likelihood, affecting the efficiency of the estimator. Moreover, the condition imposed on the covariates rules out the use of time dummies. The authors also provide identification for $T \geq 3$ and more than one lag of the dependent variable.

As shown above, conditional inference in the DL model leads to restrictive conditions on the covariates for the identification. In order to overcome this shortcomings, Bartolucci and Nigro (2010) proposed an approximation based on a model called Quadratic Exponential (QE) model, based on the multivariate binary data distribution showed by Cox (1972a). A similar approach was proposed by Bartolucci and Pennoni (2007) for the two-parameter logistic model. The QE model directly defines the conditional probability for $\boldsymbol{y}_i$ as

$$p(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp[y_{i+}\delta_i + (\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + y_{iT}(\phi + \boldsymbol{x}_{iT}'\boldsymbol{\eta}_2) + y_{i*}\psi]}{\sum_{\boldsymbol{z}} \exp[z_+\delta_i + (\sum_{t=1}^T z_t\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + z_T(\phi + \boldsymbol{x}_{iT}'\boldsymbol{\eta}_2) + z_{i*}\psi]}, \tag{1.17}$$

where the notation for parameters is different in order to distinguish them from the DL model, such that δ_i denotes the unobserved heterogeneity, $\boldsymbol{\eta}_1$ is a vector of parameters related to the set of the strictly exogenous regressors, ϕ and $\boldsymbol{\eta}_2$ are nuisance parameters,[3] and ψ denotes the state dependence. The denominator is given by the sum of all possible binary response vector $\boldsymbol{z} = (z_1, ..., z_T)$, where $z_+ = \sum_{t=1}^T z_t$ and $z_{i*} = y_{i0}z_1 + \sum_{t>1} z_{t-1}z_t$. The QE and the DL share many properties. First of all the static versions of the models coincide, namely when $\gamma = \psi = 0$. Moreover, in both specifications the state dependence parameters can be interpreted as the conditional log-odds ratio for $(y_{i,t-1}, y_t)$. Most importantly, the QE admits a sufficient statistic for the incidental parameters, namely the total score y_{i+}. Conditioning the

[3]See Bartolucci and Nigro (2010) for the discussion on parameter interpretation and additional details.

probability in Equation (1.17) on the total score leads to

$$p(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp[(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + y_{iT}(\phi + \boldsymbol{x}'_{iT}\boldsymbol{\eta}_2) + y_{i*}\psi]}{\sum_{\boldsymbol{z}(y_{i+})}\exp[(\sum_{t=1}^T z_t\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + z_T(\phi + \boldsymbol{x}'_{iT}\boldsymbol{\eta}_2) + z_{i*}\psi]}, \tag{1.18}$$

which does not depend on the incidental parameters δ_i, where $\sum_{\boldsymbol{z}(y_{i+})}$ denotes the sum over all the possible vector $\boldsymbol{z}$ of length T such that $z_+ = y_{i+}$. Consistent estimators of parameters $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \phi, \psi)$ can be obtained via the maximisation of a conditional likelihood function built summing the individual probabilities in Equation (1.18). Moreover, the estimator has a rate of convergence of $\sqrt{n}$ and is asymptotically normal. The model specification is also more flexible than those provided by previous contributions, since it allows for time dummies and it works for $T \geq 2$ beyond the initial observation.

An interesting feature of the QE model is given by the fact that it can be exploited as an approximation in order to estimate the parameter of a DL model, as argued by Bartolucci and Nigro (2012), who derive a Pseudo Conditional Maximum Likelihood estimator (PCML). The starting point is the log-probability of the DL in Equation (1.16) given by

$$\log p(\boldsymbol{y}_i|\alpha_i, y_{i0}, \boldsymbol{X}_i) = \\ y_{i+}\alpha_i + (\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta} + y_{i*}\gamma - \sum_{t=1}^T \log[1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1})]. \tag{1.19}$$

The non-linear component in Equation (1.19) is approximated by a first-order Taylor's expansion around $\alpha_i = \bar{\alpha}_i$, $\beta = \bar{\beta}$, $\gamma = 0$ as follows

$$\sum_{t=1}^T \log[1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1})] \approx \sum_{t=1}^T \{\log[1 + \exp(\bar{\alpha}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\beta}})] + \\ + \bar{q}_{i1}[\alpha_i - \bar{\alpha}_i + \boldsymbol{x}'_{it}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})]\} + \bar{q}_{i1}y_{i0}\gamma + \sum_{t>1} \bar{q}_{it}y_{i,t-1}\gamma,$$

where $\bar{\alpha}_i$ and $\bar{\boldsymbol{\beta}}$ are given values for α_i and $\boldsymbol{\beta}$ and

$$\bar{q}_{it} = \exp(\bar{\alpha}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\beta}})/[1 + \exp(\bar{\alpha}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\beta}})],$$

namely a static logit formulation for $p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it})$ at the given value of the parameters. Therefore, replacing the non-linear term with its expansion in Equation (1.19) and restoring the exponential form leads to the approximated probability given by

$$p^*(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp[y_{i+}\alpha_i + (\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta} + y_{i*}\gamma - \sum_{t=1}^T \bar{q}_{it} y_{i,t-1}\gamma]}{\sum_{\boldsymbol{z}} \exp[z_+\alpha_i + (\sum_{t=1}^T z_t\boldsymbol{x}_{it})'\boldsymbol{\beta} + z_{i*}\gamma - \sum_{t=1}^T \bar{q}_{it} z_{i,t-1}\gamma]}.$$

The last equation corresponds to a modified version of the QE model, which can be exploited for the estimation of the parameters of the DL model. As shown above, the QE model admits the total score as a sufficient statistic for parameters α_i (denoted δ_i in the QE parametrisation). Hence, by conditioning $p^*(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0})$ on the sufficient statistic y_{i+}, we obtain

$$p^*(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp[(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta} + y_{i*}\gamma - \sum_{t=1}^T \bar{q}_{it} y_{i,t-1}\gamma]}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp[(\sum_{t=1}^T z_t\boldsymbol{x}_{it})'\boldsymbol{\beta} + z_{i*}\gamma - \sum_{t=1}^T \bar{q}_{it} z_{i,t-1}\gamma]}, \tag{1.20}$$

which is independent of α_i. Finally, the probability in Equation (1.20) enters the likelihood function and the estimation procedure involves two different steps:

1. The values for $\tilde{\boldsymbol{\beta}} = (\bar{\alpha}_i, \bar{\boldsymbol{\beta}})$ are obtained via preliminary estimation of the corresponding static conditional logit model (Chamberlain, 1980). Since the estimation provides only the values of $\bar{\boldsymbol{\beta}}$, $\bar{\alpha}_i$ are computed by maximising the individual log-likelihoods of the static model, reported in Equation (1.11), evaluated in $\bar{\boldsymbol{\beta}}$, $\ell_i(\bar{\boldsymbol{\beta}})$, w.r.t. the parameter α_i.

2. The conditional log-likelihood, given the preliminary estimates, is maximised w.r.t. the set of parameters $\boldsymbol{\theta}$ and is

$$\ell^*(\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}}) = \sum_{i=1}^n \mathbb{1}(0 < y_{i+} < T) \log[p^*_{\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}}}(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+})],$$

where $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \gamma)$ and the subscript $\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}}$ denotes the fact that we are taking into account the preliminary estimates of $\bar{\alpha}_i$ and $\bar{\boldsymbol{\beta}}$.

Asymptotic properties of PCML estimator exhibits some peculiarities discussed in Bartolucci and Nigro (2012). The proposed estimator is proven to be consistent for the parameters $\boldsymbol{\beta}$ when $\gamma = 0$. Moreover, the PCML estimator results biased for the DL parameters and its bias is proportional to the magnitude of the state dependence, γ. However, simulation results suggest that PCML provides a good approximation of the DL parameters.

1.2.2 Random-effects approach

A second important methodology that allows the inclusion of heterogeneity across individuals is the random-effects approach. Individual specific effects α_i are treated as random draws that enter Equation (1.1) under some suitable distributional assumptions.

First, $\mathrm{E}(\alpha_i|\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT}) = 0$, so that the expected value equals 0 conditional on the covariates. This is a crucial point since the strong hypothesis of independence between α_i and the covariates $\boldsymbol{x}_{it}$ is imposed. It follows that this assumption often results implausible in empirical applications.

Secondly, $\mathrm{V}(\alpha_i|\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT}) = \sigma_\alpha^2$, individual effects have finite and fixed variance σ_α^2 conditional on $\boldsymbol{x}_{it}$. Finally, $\mathrm{E}(\alpha_i \varepsilon_{it}) = 0$ for $t = 1, ..., T$, requiring absence of correlation between error and the random effect.

Given the assumptions and a distribution $G(\cdot)$ for α_i, generally assumed to be a Gaussian distribution, $\alpha_i \sim N(0, \sigma_\alpha^2)$, the log-likelihood $\ell_i(\cdot)$ for each individual becomes:

$$\ell_i(\boldsymbol{\beta}, \sigma_\alpha^2) = \log \int_{-\infty}^{+\infty} \prod_{t=1}^{T} p(y_{it}|\alpha_i, \boldsymbol{x}_{it}) dG(\alpha_i). \tag{1.21}$$

In this way, individual effects are integrated out and the probability function is marginalised with respect to α_i. Since the integral does not admit a closed form solution, it is computed numerically by quadrature methods as proposed by Butler and Moffitt (1982).

Maximising the log-likelihood, $\ell(\boldsymbol{\beta}, \sigma_\alpha^2) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\beta}, \sigma_\alpha^2)$, gives a consistent estimator of parameters $\boldsymbol{\beta}$ and σ_α^2 as $n \to \infty, \forall\, T$ if the distribution $G(\cdot)$ is correctly specified. In the case of misspecification of the distribution $G(\cdot)$, consistency also requires that $T \to \infty$ (Arellano and Bonhomme, 2009).

In order to mitigate the assumption of independence between the unobserved heterogeneity and the set of covariates, Correlated Random Effects (CRE) result being a more flexible specification since they allow for correlation between α_i and the set of independent variables $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})$. The structure of dependence between individual effects and covariates is modeled with a fully parametric approach. It is required to set the conditional distribution of α_i given $\boldsymbol{X}_i$, $G(\alpha_i|\boldsymbol{X}_i)$. In this regard, two different approaches have been proposed. Mundlak (1978) suggested a linear index based on the sample average of the time-varying covariates for the i-th subject such that

$$\alpha_i = \bar{\boldsymbol{x}}_i' \boldsymbol{\lambda} + \xi_i,$$

where $\bar{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}$ is the mean over time of the covariates and $\boldsymbol{\lambda}$ is a conformable vector of parameters to be estimated. In a similar way, Chamberlain (1980) proposed a linear regression function such that

$$\alpha_i = \sum_{t=1}^{T} \boldsymbol{x}_{it}' \boldsymbol{\psi}_t + \xi_i,$$

where $\boldsymbol{\psi}_t$ is a vector of parameters and ξ_i is, in both specifications, the residual. The idiosyncratic term ξ_i is assumed to be independent of $\boldsymbol{X}_i$ and to follow a specific distribution, generally assumed $\xi_i \sim N(0, \sigma_\alpha^2)$, which will be exploited to integrate out individual effects in the likelihood function in Equation (1.21).

Further technical aspects involve dynamic models. The i-th contribution to the likelihood function can be written as a sequential factorisation as:

$$\ell_i(\boldsymbol{\beta}, \sigma_\alpha^2) = \int_{-\infty}^{+\infty} \prod_{t=0}^{T} p(y_{it}|y_{i,t-1}, \boldsymbol{x}_{it}, \alpha_i) dG(\alpha_i). \tag{1.22}$$

Due to the recursive representation, the conditional density for the first observation, y_{i0}, is not observed. Moreover, y_{i0} cannot be treated as exogenous since it depends on the parameter α_i by construction, even though this source of endogeneity becomes negligible when $T \to \infty$ (Hsiao, 2014). The underlying idea of the proposed solutions for the initial conditions is to include in the Equation (1.22) the marginal or the conditional probability of y_{i0} so that the process can be initialised.

Heckman (1981b) proposed to approximate the distribution of $y_{i0}|\alpha_i$ via a reduced form model as

$$y_{i0} = \mathbb{1}\{\theta\alpha_i + \boldsymbol{x}_{i0}'\boldsymbol{\pi} + \varepsilon_{i0}\},$$

where θ and $\boldsymbol{\pi}$ are nuisance parameters to be estimated.

Another approach is given by Wooldridge (2005) who deals with the initial condition problem in a different way. In this case the joint probability for $(y_{i1}, ..., y_{iT})$ is conditioned on the initial observation y_{i0}. It can be done by specifying a model for α_i conditional on y_{i0} and a set of strictly exogenous explanatory variables, $\boldsymbol{z}_i$, similar to the correlated random effects:

$$\alpha_i = \boldsymbol{z}_i'\boldsymbol{\pi} + y_{i0}\delta + \xi_i,$$

where $\boldsymbol{\pi}$ and δ are nuisance parameters and ξ_i is the idiosyncratic error component.

Akay (2012) provides a simulation comparison between the two different methods showing that the first has superior finite-sample performance when the panel has moderate length ($T \leq 8$), while both methods tend to perform equally for longer time dimensions.

The random-effects approach represents a convenient way to overcome the problem of inconsistency of the ML estimator due to the presence of unobserved heterogeneity. However, this approach has some shortcomings as the estimator requires proper distributional assumptions on the individual effects and the independence between α_i and $\boldsymbol{X}_i$. Though CRE partially relaxes the independence assumption, any parametrisation must involve strictly exogenous variables excluding feedback effects. Despite the general applicability, these restrictive conditions have affected the development of the random-effects approach in the most recent literature in favour of the fixed-effects approach.

1.3 Simulation study

This section proposes a Monte Carlo simulation concerning the estimation of the parameters of a DL model with fixed effects. The aim of the study is to test the finite sample performance of a set of estimators designed in order to overcome the incidental parameters problem.

1.3.1 Simulation design

The simulation design aims to generate data from a DL model by

$$y_{i0} = \mathbb{1}\{\alpha_i + \beta x_{i0} + \varepsilon_{i0} > 0\}, \tag{1.23}$$

$$y_{it} = \mathbb{1}\{\alpha_i + \beta x_{it} + \gamma y_{i,t-1} + \varepsilon_{it} > 0\}, \tag{1.24}$$

for $i = 1 \dots n$ and $t = 1 \dots T$ beyond an initial observation, in $t = 0$. Moreover, y_{it} is the binary outcome variable, $\mathbb{1}\{\cdot\}$ is the indicator function, x_{it} is an exogenous regressor generated from a Gaussian distribution with zero mean and variance $\pi^2/3$ and ε_{it} is a random variable following a logistic distribution. The parameter β is set equal to 1 and the state dependence parameter γ assumes values in $\{0.25, 0.5, 1, 2\}$ in order to evaluate different values of persistency. Individual intercepts α_i are generated as in Honoré and Kyriazidou (2000), so that $\alpha_i = \frac{1}{4}\sum_{t=0}^{3} x_{it}$. Finally, the sample sizes considered are $n = 250, 500, 1000$ and $T = 3, 4, 6, 8, 12$. The Monte Carlo replications are 1000.

In this regard, we consider estimators based on both the conditional approach, Honoré and Kyriazidou (2000) (HK)[4] and Bartolucci and Nigro (2012) (PCML), and the target-corrections proposed by Carro (2007) (MML), Dhaene and Jochmans (2015) (SPJ) and Bartolucci et al. (2016) (MPL). Moreover, the five techniques mentioned above are compared to the

[4]The bandwidth parameter is set to $\sigma_n = 8 \cdot n^{-1/5}$ in order to compare our findings with the large part of results in the original paper.

Maximum Likelihood estimator (ML) and to the Infeasible Maximum Likelihood estimator (INF), where the true values of the individual intercepts are included in the model as an additional regressor and the set of parameters is estimated by maximum likelihood. Dhaene and Jochmans (2015) proposed a simulation study with the same design, where they compare a wide set of target-corrected estimators. In this study, we focus on the most recent methodological contributions in that field (i.e. MML,SPJ,MPL) and the the two conditional estimators.

1.3.2 Simulation results

This section reports the main results of the simulation study. As in the original contribution of Honoré and Kyriazidou (2000), the true value of the state dependence parameter is set to $\gamma = 0.5$ for the benchmark design.[5] Table 1.1 and Table 1.2 show the statistics of the seven considered estimators for the parameters β and γ, respectively. For each sample size, the mean bias, the median bias, the root mean square error and the median absolute error are reported.

First of all, we can observe that the incidental parameters problem has an high impact on the bias of the ML estimator. As expected, the bias is considerable regardless the sample size and it slightly tends to decrease as the time series length grows. Moreover, the bias appears to be larger for the estimator of the state dependence parameter, $\hat{\gamma}$. On the contrary, the infeasible likelihood estimator, INF, performs well[6] and its finite sample bias is always negligible for the whole set of parameters.

The behaviour of the five estimators showed above is not homogeneous. As theory would suggest, the target corrected estimators are sensitive to the number of observations over time because the larger T, the lower the magnitude of the bias. This can be easily verified in the tables. For a given n, the bias of MML, MPL, and SPJ[7] shrinks as the time series grows,

[5] The full set of simulation results are reported in Appendix A.

[6] By construction, the INF estimator is not affected by the incidental parameters problem.

[7] The SPJ estimator is computed for $T \geq 6$.

Table 1.1: Simulation results under Benchmark design based on Equations (1.23) and (1.24) with $\gamma = 0.5$, $\beta = 1$

Results for $\hat{\beta}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.007	0.817	0.149	0.027	0.002	-0.050	-	0.087	0.865	0.375	0.145	0.179	0.090	-
	4	0.010	0.572	0.057	0.015	0.013	0.009	-	0.074	0.600	0.165	0.102	0.080	0.077	-
	6	0.008	0.327	0.029	0.009	0.027	0.025	-0.152	0.057	0.343	0.102	0.069	0.070	0.070	0.209
	8	0.008	0.224	0.017	0.007	0.021	0.020	-0.121	0.048	0.237	0.071	0.059	0.061	0.061	0.153
	12	0.004	0.134	0.010	0.004	0.010	0.010	-0.057	0.041	0.144	0.052	0.045	0.046	0.046	0.077
500	3	0.002	0.777	0.069	0.006	-0.059	-0.055	-	0.064	0.800	0.200	0.095	0.106	0.075	-
	4	0.004	0.552	0.031	0.005	0.007	0.004	-	0.051	0.566	0.110	0.069	0.055	0.054	-
	6	0.001	0.316	0.015	0.002	0.020	0.019	-0.140	0.041	0.324	0.068	0.048	0.050	0.050	0.172
	8	0.001	0.216	0.010	0.002	0.015	0.014	-0.117	0.033	0.222	0.049	0.038	0.040	0.040	0.131
	12	0.003	0.131	0.006	0.002	0.008	0.008	-0.057	0.026	0.135	0.035	0.029	0.030	0.030	0.067
1000	3	0.002	0.764	0.046	0.002	-0.019	-0.056	-	0.044	0.776	0.138	0.066	0.095	0.067	-
	4	0.002	0.546	0.024	0.002	0.005	0.002	-	0.035	0.553	0.082	0.048	0.038	0.037	-
	6	0.001	0.314	0.011	0.001	0.019	0.018	-0.135	0.029	0.319	0.050	0.035	0.038	0.038	0.151
	8	0.002	0.215	0.008	0.001	0.014	0.014	-0.112	0.024	0.218	0.037	0.028	0.031	0.031	0.120
	12	0.000	0.129	0.005	0.001	0.007	0.007	-0.054	0.019	0.132	0.025	0.020	0.022	0.021	0.059

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.006	0.782	0.078	0.009	-0.039	-0.054	-	0.061	0.782	0.168	0.090	0.072	0.066	-
	4	0.003	0.560	0.044	0.007	0.010	0.007	-	0.047	0.560	0.101	0.065	0.051	0.051	-
	6	0.005	0.319	0.021	0.006	0.023	0.022	-0.145	0.035	0.319	0.062	0.045	0.049	0.048	0.148
	8	0.006	0.217	0.019	0.006	0.021	0.020	-0.119	0.037	0.217	0.048	0.039	0.043	0.042	0.120
	12	0.006	0.134	0.011	0.005	0.011	0.011	-0.058	0.029	0.134	0.036	0.030	0.031	0.031	0.060
500	3	0.000	0.759	0.041	0.001	-0.055	-0.056	-	0.041	0.759	0.115	0.060	0.063	0.057	-
	4	0.003	0.545	0.025	0.002	0.006	0.003	-	0.034	0.545	0.071	0.045	0.037	0.036	-
	6	0.001	0.314	0.013	0.001	0.020	0.019	-0.139	0.027	0.314	0.044	0.032	0.034	0.034	0.140
	8	-0.000	0.214	0.008	0.000	0.014	0.013	-0.116	0.021	0.214	0.033	0.026	0.027	0.027	0.116
	12	0.003	0.131	0.004	0.003	0.009	0.009	-0.058	0.017	0.131	0.023	0.020	0.021	0.021	0.058
1000	3	-0.001	0.750	0.028	-0.003	-0.043	-0.057	-	0.030	0.750	0.079	0.043	0.054	0.057	-
	4	0.003	0.539	0.022	-0.000	0.003	-0.000	-	0.024	0.539	0.052	0.032	0.025	0.025	-
	6	0.001	0.312	0.009	-0.000	0.018	0.017	-0.133	0.020	0.312	0.032	0.022	0.025	0.024	0.133
	8	0.001	0.214	0.006	0.001	0.015	0.013	-0.112	0.017	0.214	0.024	0.019	0.021	0.021	0.112
	12	0.000	0.130	0.005	0.001	0.007	0.006	-0.054	0.013	0.130	0.017	0.014	0.014	0.014	0.054

Table 1.2: Simulation results under Benchmark design based on Equations (1.23) and (1.24) with $\gamma = 0.5$, $\beta = 1$

Results for $\hat{\gamma}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	-0.007	-2.617	-0.053	-0.000	-0.202	-0.571	-	0.149	2.690	0.690	0.438	0.370	0.620	-
	4	-0.000	-1.576	-0.036	0.017	-0.208	-0.318	-	0.126	1.615	0.355	0.280	0.302	0.382	-
	6	-0.003	-0.907	-0.058	0.003	-0.110	-0.142	0.770	0.103	0.933	0.232	0.190	0.210	0.224	0.845
	8	-0.001	-0.649	-0.080	-0.012	-0.074	-0.087	0.286	0.094	0.669	0.182	0.150	0.165	0.168	0.357
	12	-0.002	-0.397	-0.057	-0.003	-0.028	-0.032	0.098	0.071	0.415	0.130	0.112	0.115	0.115	0.162
500	3	-0.000	-2.588	-0.058	0.003	-0.229	-0.567	-	0.107	2.621	0.397	0.288	0.293	0.590	-
	4	-0.002	-1.580	-0.051	0.009	-0.214	-0.324	-	0.089	1.600	0.257	0.199	0.263	0.357	-
	6	-0.001	-0.909	-0.050	0.001	-0.114	-0.145	0.743	0.070	0.921	0.166	0.133	0.168	0.188	0.782
	8	-0.000	-0.630	-0.041	0.006	-0.059	-0.070	0.298	0.058	0.640	0.126	0.104	0.116	0.122	0.330
	12	0.004	-0.387	-0.052	0.006	-0.019	-0.023	0.104	0.049	0.395	0.097	0.077	0.078	0.079	0.140
1000	3	0.002	-2.548	-0.031	0.021	-0.233	-0.559	-	0.075	2.563	0.273	0.197	0.286	0.570	-
	4	-0.001	-1.577	-0.052	0.009	-0.217	-0.324	-	0.065	1.587	0.193	0.142	0.245	0.342	-
	6	-0.002	-0.901	-0.043	0.007	-0.109	-0.138	0.740	0.051	0.907	0.113	0.090	0.138	0.160	0.758
	8	0.002	-0.628	-0.038	0.006	-0.058	-0.070	0.288	0.041	0.634	0.093	0.074	0.092	0.099	0.305
	12	-0.001	-0.397	-0.046	-0.004	-0.030	-0.033	0.091	0.034	0.402	0.079	0.058	0.064	0.065	0.111
		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	-0.010	-2.563	-0.050	0.015	-0.197	-0.561	-	0.108	2.563	0.394	0.300	0.233	0.561	-
	4	-0.000	-1.570	-0.061	-0.000	-0.211	-0.327	-	0.082	1.570	0.253	0.189	0.228	0.327	-
	6	-0.007	-0.921	-0.058	-0.012	-0.121	-0.161	0.760	0.072	0.921	0.154	0.130	0.157	0.179	0.760
	8	-0.001	-0.650	-0.088	-0.006	-0.079	-0.083	0.287	0.071	0.650	0.128	0.109	0.102	0.107	0.287
	12	-0.002	-0.400	-0.054	-0.004	-0.030	-0.033	0.097	0.048	0.400	0.089	0.078	0.079	0.078	0.112
500	3	-0.000	-2.586	-0.054	0.000	-0.225	-0.564	-	0.073	2.586	0.259	0.199	0.227	0.564	-
	4	-0.000	-1.575	-0.054	0.011	-0.215	-0.325	-	0.058	1.575	0.177	0.131	0.219	0.325	-
	6	0.000	-0.906	-0.058	0.005	-0.111	-0.143	0.732	0.046	0.906	0.114	0.097	0.118	0.144	0.732
	8	-0.001	-0.629	-0.041	0.007	-0.059	-0.069	0.293	0.038	0.629	0.087	0.071	0.081	0.084	0.293
	12	0.007	-0.392	-0.060	-0.002	-0.028	-0.030	0.102	0.028	0.392	0.075	0.048	0.054	0.055	0.104
1000	3	0.001	-2.530	-0.033	0.022	-0.227	-0.558	-	0.050	2.530	0.180	0.130	0.228	0.558	-
	4	0.001	-1.571	-0.051	0.012	-0.215	-0.321	-	0.042	1.571	0.132	0.094	0.215	0.321	-
	6	-0.003	-0.899	-0.044	0.008	-0.105	-0.136	0.734	0.035	0.899	0.078	0.062	0.108	0.136	0.734
	8	0.002	-0.630	-0.040	0.005	-0.058	-0.070	0.286	0.028	0.630	0.062	0.051	0.066	0.074	0.286
	12	0.001	-0.400	-0.050	-0.006	-0.031	-0.035	0.088	0.022	0.400	0.058	0.039	0.045	0.047	0.089

even though the split panel jackknife procedure requires $T \geq 8$ to produce a remarkable bias reduction. Among this class of estimators, the MPL performs better in the two smallest configuration length considered.

On the other hand, we have the conditional approach. The bias of the HK estimator is small for both parameters. The bias of $\hat{\beta}$ tends to reduce as both n and T increase, while the bias of $\hat{\gamma}$ is stable for each configuration even though larger sample sizes reduce the median absolute error. This fact is due to the choice of the parameter of the kernel function. Finally, the PCML estimator shows the best performance. The bias is the smallest in almost the whole set of configurations. Clearly, given the identification strategy, larger values of n and T lead the root mean square error and the median absolute error to decrease. To sum up, for this design, the conditional approach exhibits remarkable relative advantage in terms of bias with respect to the the target-adjusted estimators when $T \leq 8$.

As pointed out by Dhaene and Jochmans (2015), as the state dependence parameter becomes larger, data are less informative for fixed effects models. In this regard, Table 1.3 reports results about the bias of the five examined estimators for different values of γ. In order to do so, we exploit the Δ index proposed in Bartolucci et al. (2016), given by

$$\Delta(*) = \frac{|MB(ML)| - |MB(*)|}{|MB(ML)| - |MB(INF)|},$$

where $|MB()|$ is the absolute value of the median bias of the estimators. This index can be seen as a relative perfomance of an estimator, $(*)$, with respect to the INF, where the ML represents a benchmark.

First of all, we note that the effect of a variation in the true value of the state dependence parameter is different for $\hat{\beta}$ and $\hat{\gamma}$. In fact, the Δ index is stable for all the five estimators of β as the true γ grows.

On the contrary, the behaviour of the estimator $\hat{\gamma}$ tends to vary according to the estimation technique. In fact, we have again to distinguish between the two approaches. Target-adjusted estimators are more sensitive to the state dependence with respect to the "conditional" estimators. Specifically, the performance of MPL and MML worsens as γ increases, while SPJ tends

Table 1.3: Simulation results under Benchmark design.
Relative Performance (Δ), $n = 500$

		$\hat{\beta}$					$\hat{\gamma}$				
γ_0	T	HK	PCML	MPL	MML	SPJ	HK	PCML	MPL	MML	SPJ
0	3	0.993	0.994	0.915	0.926	-	0.986	1.003	1.004	0.860	-
	6	1.002	1.005	0.959	0.957	0.568	1.001	0.999	0.912	0.887	0.195
	12	1.022	1.014	0.974	0.975	0.621	1.016	0.992	0.996	0.985	0.686
0.25	3	0.903	0.998	0.919	0.924	-	0.970	0.998	0.972	0.835	-
	6	0.980	1.003	0.954	0.955	0.504	0.959	0.999	0.893	0.869	0.212
	12	0.965	0.995	0.949	0.953	0.581	0.911	1.000	0.941	0.939	0.746
0.5	3	0.947	1.000	0.928	0.927	-	0.979	1.000	0.913	0.782	-
	6	0.961	1.000	0.938	0.942	0.557	0.936	0.995	0.878	0.842	0.192
	12	0.996	1.001	0.953	0.957	0.571	0.862	1.012	0.946	0.939	0.753
1	3	0.890	1.005	0.954	0.937	-	0.961	0.985	0.819	0.691	-
	6	0.957	1.023	0.953	0.957	0.629	0.898	1.004	0.840	0.787	0.219
	12	0.961	1.002	0.952	0.962	0.596	0.737	0.997	0.914	0.899	0.821
2	3	0.829	1.006	0.975	0.934	-	0.939	0.915	0.614	0.467	-
	6	0.930	0.993	0.924	0.944	0.649	0.782	0.991	0.700	0.615	0.207
	12	0.982	0.978	0.935	0.950	0.662	0.492	0.977	0.858	0.817	0.889

to be more robust even though it shows a bigger order of magnitude of the bias.

Finally, a peculiar point concerns the PCML estimator. It is worth recalling that it is derived by an approximation, so we could expect that its relative performance is decreasing in γ, since the PCML is proven to be consistent only for $\gamma = 0$. The table shows that not only the Δ index reported in Table 1.3 is stable, but its is also the best estimator for large values of the state dependence. This result is due to the small bias of the estimator and is in line with the findings of Bartolucci and Nigro (2012) where the bias of the PCML is reported for a set of simulations with different levels of state dependence.

1.4 Conclusions

This work reviews estimation of panel data binary choice models. The inclusion of individual unobservable characteristics and the role of dynamics are crucial points for applied analyses. However, dealing with heterogeneity leads to the failure of the standard maximum likelihood framework.

The random-effects approach is a simple and general solution based on many distributional assumptions. The correlated random-effects approach relaxes some of those assumptions making this methodology more appealing for practitioners. The main advantage of this approach is that it is possible to identify the effects of time-invariant explanatory variables.

On the other hand, the fixed-effects approach literature has been facing a remarkable expansion. Two different methodologies try to overcome the incidental parameters problem.

Target-corrected estimators are a wide class of techniques that produce a reduced bias estimates. The main advantage is given the generality of the approach. The conditional approach exploits parameters identification strategies and sufficient statistics in order to obtain estimates independently from the incidental parameters, even though it is model specific.

Finally, a simulation study investigates finite sample properties of different seven estimators for the parameters estimation of a DL model. The experiment shows that the conditional estimators perform extremely better than the bias-corrections when the panel length is small. However, as the number of observations over time becomes larger, they tend to produce equivalent results.

Chapter 2

A conditional approach for testing endogenous sample and self-selection in fixed-effects logit panel data models

2.1 Introduction

The use of binary choice non-linear models is a standard practice in panel data econometrics. However, practitioners have to face well known issues in applied works: a potential endogenous binary explanatory variable, also defined as self-selection, or the fact that some observations of a response variable are not missing at random (Little and Rubin, 2019), that is sample selection, namely a potential mechanism that drives missingness. An economist could be interested in testing the endogeneity of self- or sample selection mechanisms, since both problems could lead to biased estimators. Although these issues have been widely discussed for cross-sectional linear models, the panel structure of the data and the non-linearity of the models here considered have to be taken into account.

As concerns the second aforementioned problem, the linear probability model could be an attractive tool. Even though it provides good approximations of covariates' marginal effects, predictions based on those estimates are not reliable. Moreover, a simple two-stages procedure where the fitted values of the endogenous explanatory variables from the first-stage are included in the second-stage non-linear equation would lead to biased estimators and it is know as the "forbidden regression" (Hausman, 1975). Furthermore, in panel data analysis dealing with unobserved heterogeneity is a crucial point. One way is to rely on parametric assumptions exploiting the well known random-effects model. However, some of these hypotheses about the distribution of unobserved heterogeneity could be unreliable in economic applications. A feasible strategy to overcome this problem is the fixed-effects approach even though standard maximum likelihood estimates are affected by the "incidental parameter problem" as argued by Neyman and Scott (1948). Moreover, as pointed out by Lin and Wooldridge (2019), it is crucial to distinguish between different sources of potential endogeneity, the one due to time-invariant unobserved components that simultaneously affect the outcome variable and the selection mechanism (*heterogeneity endogeneity*) and the correlation of idiosyncratic shocks, which is defined *idiosyncratic endogeneity.*[1]. In this regard, Semykina and Wooldridge (2018) propose a strategy that allows the estimation of a probit model with a selection equation under the assumption of normally distributed errors and a simple test for the presence of endogenous sample selection. The issue of heterogeneity endogeneity is dealt assuming the relation between unobservables and covariates by the Correlated Random-Effects (CRE) approach (Mundlak, 1978).

The present work introduces a testing procedure for the endogeneity of a self- or sample selection mechanism in binary choice panel data models with fixed effects. This methodology relies on an approximation of a fixed-effects logit model formulation that can be easily estimated by conditional maximum likelihood in a two-step procedure and that admits a very simple variable-addition test. Since this approach is of a fixed-effects nature, the individual unobserved heterogeneity is treated non-parametrically, as opposed to the CRE approach, which requires assuming the functional form for the correlation between the individual effects and the model covariates. More-

[1] Hereafter, we will refer to the latter as source of "endogeneity"

over, a Pseudo Conditional Maximum Likelihood (PCML) estimator exploits sufficient statistics for the individual effects, which are therefore allowed to be arbitrarily correlated with the regressors and among each other. The main advantage of the proposed test is that it is always able to identify the idiosyncratic endogeneity since the choice of the "conditional inference" approach allows to handle heterogeneity endogeneity independently and to overcome the incidental parameters problem at the same time. An extensive Monte Carlo study based on different designs is presented in order to show the finite sample performance of the proposed testing procedure and to compare it with the existing alternatives. This experiment shows that the proposed methodology is more robust to different data generating processes because of it does not require parametric assumptions about the unobserved heterogeneity. Finally, the work proposes an application to real data about the health conditions of a set of individuals where we consider the retirement status being a potential endogenous variable. Coherently with previous empirical applications, the results suggest that the retirement status is endogenous and it has a negative and significant impact of retirement on the probability of being in bad health conditions.

The rest of the paper is organized as follows: a brief literature review is reported in Section 2.2. Section 2.3 describes the proposed model formulation. The pseudo conditional likelihood estimator is presented in Section 2.4. The proposed test procedure and the existing alternatives provided by the literature are described in Section 2.5. Section 2.6 shows how to extend the previous results to dynamic models. In Section 2.7 we show the simulation design and discuss the main simulation results concerning the test. Finally, in Section 2.8 we find applications on real data and Section 2.9 concludes.

2.2 Literature Review

This section discusses the main contributions in the related literature. In particular, Section 2.2.1 summarizes the main theoretical aspects concerning the estimation of binary choice models with fixed-effects reported in Section 1.2.1, while Section 2.2.2 and Section 2.2.3 report contributions concerning endogenous binary explanatory variables and sample selection, respectively.

2.2.1 Estimation of fixed-effects binary choice panel data models

This section recalls the main points of the contributions examined in Section 1.2.1. The literature concerning the fixed-effects approach to the estimation of binary choice panel data models includes a wide variety of techniques. Standard maximum likelihood estimator is proved to be biased because of the incidental parameter problem, as argued by Neyman and Scott (1948) and Lancaster (2000). Among the different proposals in order to overcome this problem, it is possible to identify two main groups: conditional inference and target-adjusted estimators.

The conditional inference approach exploits sufficient statistics for the incidental parameters and derives conditional likelihood functions for the estimation of the structural parameters. Seminal contributions in this field are from Andersen (1970), McFadden (1974), and Chamberlain (1980), who derive the conditional static logit model and the conditional maximum likelihood estimator that allows us to consistently estimate the parameters of the model, regardless the nuisance parameters. The extension of the conditional inference results to the dynamic logit (DL) model has not been straightforward. Chamberlain (1993) describes the conditions for a conditional approach estimation of the DL model without regressors, while the estimator proposed by Honoré and Kyriazidou (2000) allows for exogenous regressors exploiting a kernel density function. A more recent contribution is the one of Bartolucci and Nigro (2010) who propose an approximation for the DL model based on the Quadratic Exponential model (QE) originally described by Cox (1972a) and a PCML estimator for the DL model parameters, based on a modified version of the QE, reported in Bartolucci and Nigro (2012).

The target-adjusted approach is a more complex scenario. However, all the estimators have the property to lower the bias of the maximum likelihood estimator (MLE). As argued by Arellano and Hahn (2007), it is possible to identify three main subgroups of techniques: (i) bias-corrected estimators,

(ii) score-corrected estimators, (iii) and the objective-function approach. The first group directly deals with the bias of the MLE. In particular, Hahn and Newey (2004) propose an analytical formula for the bias in nonlinear static panel data models. The extension to dynamic models has been proposed by Hahn and Kuersteiner (2011), while Fernández-Val (2009) focuses on the specific analytical formula of the bias for binary choice models. Hahn and Newey (2004) also introduce a jackknife for static models and Dhaene and Jochmans (2015) deal with the dynamic version of the jackknife based on split-panel. Moreover, an extension of the analytical corrections and the split-panel jackknife for individual and time effects is provided by Fernández-Val and Weidner (2016). Concerning the score-corrected approach, Carro (2007) exploits a modified estimating equation that leads to a modified maximum likelihood estimator. Finally, the objective function approach is based on modifications of the likelihood function. Bester and Hansen (2009) and Arellano and Hahn (2016) define penalty functions and propose to maximize the penalized likelihood function. Similarly, Bartolucci et al. (2016) propose a modified profile likelihood approach.

2.2.2 Endogenous explanatory variables

The issue of endogenous explanatory variables has been widely discussed for linear models, where standard techniques have been developed such as the two-step least squares Instrumental Variables (IV) or the Control Function (CF) approaches (Greene, 2011). A generalization for non-linear models of the IV and the CF techniques is provided in Terza et al. (2008), who propose a simple two-step estimation procedure that mimics the CF approach, defined as *two-step residuals inclusion*. Moreover, a variety of non-parametric and semiparamteric estimators are reported in Blundell and Powell (2003).

The treatment of endogenous explanatory variables is different when we are interested in the estimation of non-linear binary choice models. Different methods to deal with continuous endogenous explanatory variables are explained in Rivers and Vuong (1988) and all rely on the joint distributional assumption of the error terms. It is worth mentioning the limited information maximum likelihood based on the joint density of the two equations, the instrumental variable probit model (Lee, 1981), the generalized two-stage simultaneous probit (Amemiya, 1978) and a minimum chi-squared estimator (Newey, 1987) that exploits reduced form for the endogenous variables. Moreover, a relevant contribution is the two-stage conditional maximum likelihood (Rivers and Vuong, 1988) that defines the joint density by a factorization of a marginal distribution for the endogenous variables and a conditional one for the binary outcome. An additional peculiarity arises given when the endogenous variable is discrete. The seminal work of Heckman (1978) proposes a general framework that allows for the presence of a dichotomous endogenous variable. The model includes two binary outcomes that are simultaneously determined where, again, the estimation is based on the assumption of multivariate normally distributed errors.

As well as the estimation of the aforementioned models, studies are devoted to test for endogeneity. Monfardini and Radice (2008) describe the variety of available tests in literature for the bivariate probit model. Inside the joint estimation in the maximum likelihood framework we have the classical Likelihood Ratio (LR) and Wald Tests. Alternatives are given by the Lagrange Multipliers test and the conditional moments test based on a univariate probit model. The LR exhibits the best performance. Despite their simplicity, tests based on univariate models also perform well. Recently, Wooldridge (2014) proposed a straightforward variable addition test for endogeneity of a dichotomous regressor in binary choice models. It consists in a two-steps CF approach in a quasi-limited information maximum likelihood framework, where the residuals obtained from a first-step estimation are used as a regressor in the second step. A wide review of the use of the control function approach for linear and non-linear models is provided by Wooldridge (2015).

Finally, Lin and Wooldridge (2017) proposed a methodology to estimate a model for a binary outcome with many endogenous explanatory variables, one of which is binary. This technique consists in estimate a bivariate probit model where a CF approach is used to take into account the endogeneity of the covariates. Also in this case authors proposed a straightforward test of exogeneity based on a two-step procedure.

The arguments above can be extended to panel data analysis. The key point is taking into account the specification of the unobserved heterogeneity. In this regard, Papke and Wooldridge (2008) propose a CRE approach, where the potential correlation between time-invariant unobservables and covariates is modeled by a linear specification as suggested by Mundlak (1978), for the estimation of fractional response variables.

Further, Lin and Wooldridge (2019) show a way to handle the two different sources of endogeneity in linear and non-linear models. The correlation between covariates and a time-invariant unobserved component is treated by the CRE, while the correlation of continuous variables and time-varying errors exploits the CF approach, which allows us to test for idiosyncratic endogeneity by a two-steps estimation procedure. However, the specification of the time-constant unobservables relies on strong parametric assumptions that could result quite restrictive in economic applications.

2.2.3 Sample Selection

Sample selection issues often occur in economic applications. The contributions of Heckman (1974, 1976, 1979) are fundamental in this field and define the problem of a non-random selected sample as an "omitted variable problem". The normality distributional assumption allows us to estimate linear models in a maximum likelihood framework or a two-step procedure. A wide set of econometric methods has been recently reviewed in Pigini (2015). Literature also provides different ways in order to test sample selection. Similar to the endogenous dummy variable case, Vella (1992) presents a conditional moment test and a t-test able to detect the presence of sample selection.

With reference to the issue of non-random selected sample in binary choice models, it is worth recalling the contribution of Van de Ven and Van Praag (1981). Authors analyzed the preference for health insurance with a deductible in Netherlands and proposed a probit model with sample selection estimated by a maximum likelihood approach. The errors of the selection and the main equation are assumed to follow a bivariate normal distribution, following the methodology of Heckman (1979) for linear models.

The estimation of econometric models based on unbalanced panel data is often threatened by sample selection issues, which may occur if the absent information cannot be considered as missing at random.

Several approaches have been proposed to tackle sample selection issues in linear panel data models with unobserved effects. A wide range of parametric and non-parametric two-step estimators has been developed (Ahn and Powell, 1993; Wooldridge, 1995; Kyriazidou, 1997; Rochina-Barrachina, 1999). Further techniques concerns dynamic models (Arellano et al., 1999; Kyriazidou, 2001; Gayle and Viauroux, 2007; Semykina and Wooldridge, 2013) and the inclusion of endogenous explanatory variables in the model specification (Vella and Verbeek, 1999; Semykina and Wooldridge, 2010).

Despite the variety of approaches available for the linear model, the methodological literature has devoted relatively little attention on how to deal with sample selection in binary choice models for unbalanced panel data. One exception is the contribution by Semykina and Wooldridge (2018), who propose estimating a probit model with binary selection equation by Maximum Likelihood. The bivariate normality assumption allows them to rely on a control function approach to account for the non-random missing information. In addition, the potential correlation between the unobserved individual effects and the model covariates is dealt with by the CRE approach. Authors also proposed a variable addition test for sample selection that relies on a two-steps estimate of the parameters.

2.3 Model formulation

The present section introduces the proposed models. Despite the statistical models for the self and the sample selection are similar, they are presented separately in order to highlight their peculiarities.

2.3.1 Endogenous binary explanatory variable

For the response variables s_{it} and y_{it}, with $i = 1, \dots, n$ and $t = 1, \dots, T$, consider the model based on assuming that

$$s_{it} = \mathbb{1}(s^*_{it} > 0), \tag{2.1}$$

$$y_{it} = \mathbb{1}(y^*_{it} > 0), \tag{2.2}$$

where Equations (2.1) and (2.2) are defined as selection and main equation, respectively. Moreover, $\mathbb{1}\{\cdot\}$ is the indicator function, and s^*_{it} and y^*_{it} are latent variables defined as

$$s^*_{it} = \eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma} + u_{it}, \tag{2.3}$$

$$y^*_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + s_{it}\delta + \varepsilon_{it}. \tag{2.4}$$

In the previous expressions, $\boldsymbol{w}_{it} = (\boldsymbol{x}'_{it}, m_{it})'$ and $\boldsymbol{x}'_{it}$ are time-varying strictly exogenous individual covariates and we assume that $\boldsymbol{w}_{it}$ includes at least a strong and exogenous exclusion restriction, denoted m_{it}. Moreover, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and δ represents the set of parameters, while α_i and η_i denotes the sets of individual time-constant unobserved characteristics that are allowed to be correlated with each other and with the model's covariates. Inside the adopted fixed-effects approach unobserved heterogeneity is modeled as individual specific intercepts. Furthermore, in order to model the relationship between the error terms in the two equations, we define

$$\varepsilon_{it} \equiv u_{it}\rho + v_{it}\sqrt{1-\rho^2}, \tag{2.5}$$

where u_{it} and v_{it} are independent error terms with standard logistic distribution and ρ is the correlation coefficient between the error terms in the two equations, respectively u_{it} and ε_{it}.

In this scenario the dichotomous variable s_{it} enters the latent variable y^*_{it}, for the outcome y_{it} as in the main Equation (2.4) and the endogeneity is defined according to the parameter ρ in Equation (2.5). Clearly, when $\rho = 0$ the correlation vanishes and we face two independent equations, implying the exogeneity of s_{it}. On the contrary, the presence of the correlation affects the conditional expected value of y_{it} given $\boldsymbol{x}_{it}$ and s_{it}. Therefore, we need to derive the distribution of y_{it} conditional on s_{it}.

In this regard, from the Bayes' rule, we have that

$$p(y_{it}|s_{it}=1,\alpha_i,\boldsymbol{x}_{it}) = \frac{p(y_{it},s_{it}=1|\alpha_i,\eta_i,\boldsymbol{x}_{it},\boldsymbol{w}_{it})}{p(s_{it}=1|\eta_i,\boldsymbol{w}_{it})},$$
$$p(y_{it}|s_{it}=0,\alpha_i,\boldsymbol{x}_{it}) = \frac{p(y_{it},s_{it}=0|\alpha_i,\eta_i,\boldsymbol{x}_{it},\boldsymbol{w}_{it})}{p(s_{it}=0|\eta_i,\boldsymbol{w}_{it})},$$

and, since u_{it} is assumed to follow a logistic distribution, then s_{it} follows the logit model

$$p(s_{it}=1|\eta_i,\boldsymbol{w}_{it}) = \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma}),$$

where $\Psi(\cdot)$ denotes the the standard logistic distribution function.

The main point is that, under the set of assumptions, we do not know the joint probability of y_{it} and s_{it} and therefore the conditional distribution of $p(y_{it}=1|s_{it})$. However, an approximated version of $p(y_{it}=1|s_{it})$ may be based on a linear approximation of $\log p(y_{it}=1|s_{it})$ similar to the one proposed by Nicoletti and Peracchi (2001). In this regard, we denote the joint probability of y_{it} and s_{it} with $p_\rho(y_{it},s_{it})$ to stress its dependence on the correlation parameter ρ. The approximation is given by a first-order Taylor[2] expansion around $\rho=0$, such that

$$\log p_\rho(y_{it},s_{it}) \approx \log p_0(y_{it},s_{it}) + \frac{\partial \log p_0(y_{it},s_{it})}{\partial\rho}\rho, \tag{2.6}$$

with

$$\frac{\partial \log p_0(y_{it},s_{it})}{\partial\rho} = \frac{1}{p_0(y_{it},s_{it})}\frac{\partial p_0(y_{it},s_{it})}{\partial\rho}.$$

Following the result[3] in Equation (2.6), the approximated conditional distributions of y_{it} given s_{it} is defined as a logit model with the latent variable of the main equation augmented by an additional regressor defined in Equation

[2]Similarly, also Bartolucci and Nigro (2012) derived an approximation for the DL model based on a first order Taylor expansion of the Quadratic Exponential Model.

[3]The full derivation of the approximation is reported in Appendix B.

(2.8). Therefore, we approximate this distribution as

$$p^*(y_{it} = 1|s_{it}) = \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + s_{it}\delta + q_{it}\rho). \tag{2.7}$$

The term q_{it} in the last equation corresponds to the conditional expected value of the error term of the selection rule in Equation (2.3). It is derived by exploiting the assumption of u_{it} logistically distributed, so that, following Arabmazar and Schmidt (1982), we have[4]

$$q_{it} = \begin{cases} q^1_{it} \text{ if } s_{it} = 1, \\ q^0_{it} \text{ if } s_{it} = 0, \end{cases} \tag{2.8}$$

with

$$q^1_{it} = \mathrm{E}(u_{it}|u_{it} > -\bar{u}_{it}) = -(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma}) - \frac{\ln[1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]}{\Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})} \tag{2.9}$$

and

$$q^0_{it} = \mathrm{E}(u_{it}|u_{it} < -\bar{u}_{it}) = -(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma}) + \frac{\ln[\Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]}{[1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]}, \tag{2.10}$$

where $\bar{u}_{it}$ denotes the linear index $\bar{u}_{it} = \eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma}$. The approximating model in Equation (2.7) links the additional regressor q_{it} to the correlation coefficient ρ and it clearly coincides with the the main equation when $\rho = 0$, which is a standard fixed effects logit model.

This result is very similar to the model proposed in Heckman (1979) in his seminal paper. Although the Heckman's proposal has been widely employed by practitioners, it was criticized by Puhani (2000) and it is appropriate to recall the main points of that critique: the parameters identification and the distributional assumptions. The identification problem arises when there are the same variables in $\boldsymbol{x}_{it}$ and $\boldsymbol{w}_{it}$. In order to overcome this issue, we can exploit the non-linearity of the correction term in Equations (2.9) and (2.10) as the only source of identification of the approximating model. Unfortunately,

[4]Arabmazar and Schmidt (1982) provide the result for $\mathrm{E}(u_{it}|u_{it} > -\bar{u}_{it})$. The formulation for $\mathrm{E}(u_{it}|u_{it} < -\bar{u}_{it})$ is given in Appendix C.

as well as the inverse Mill's ratio, the correction term tends to be a linear function in some points of its domain and this fact could lead to collinearity. A suitable solution is to include a set of exclusion restrictions in $\boldsymbol{w}_{it}$. However, appropriate variables are often not available in empirical applications and this aspect could lead to a weak identification problem. Secondly, the model formulation includes the assumption of logistically distributed errors. Puhani (2000) pointed out that estimated coefficients could be sensitive to potential distributional misspecifications, in the selection and the main equation. Differently from the linear models, the distributional assumptions are crucial since their misspecification could lead to an inconsistent estimator.

2.3.2 Sample selection

The framework presented in the previous section can be adapted to tackle the issue of the sample selection. Define now a binary choice model for the response y_{it}, subject to a selection rule driven by a binary selection variable, s_{it}. For a sample of n units observed over T time occasions, the model with sample selection is based on the following observational rule:

$$s_{it} = \mathbb{1}(s_{it}^* > 0), \tag{2.11}$$

$$y_{it} = \mathbb{1}(y_{it}^* > 0) \text{ if } s_{it} = 1, \text{ not observed otherwise}, \tag{2.12}$$

where the latent variables s_{it}^* and y_{it}^* are now defined as

$$s_{it}^* = \eta_i + \boldsymbol{w}_{it}'\boldsymbol{\gamma} + u_{it}, \tag{2.13}$$

$$y_{it}^* = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \tag{2.14}$$

As well as in the model for self-selection, we follow the same set of assumptions defined in the previous section for the Equations (2.3), (2.4), and (2.5), with the only difference that in this case s_{it} is no longer an explanatory variable in the main equation but it is a sample-selection variable.

As in the previous case, this model admits correlation between the two error terms. When $\rho = 0$ the two equations are independent, so we can consider observations missing at random. However, when the correlation is non-zero, ($\rho \neq 0$), the selection process affects the conditional expected value of y_{it} given $\boldsymbol{x}_{it}$ and therefore we have to consider the distribution of y_{it} conditional on the fact of being observed ($s_{it} = 1$). From the statistical point of view, this situation partially coincides with the problem shown in Section 2.3.1. Again, from the Bayes' rule, we have

$$p(y_{it}|s_{it} = 1, \alpha_i, \boldsymbol{x}_{it}) = \frac{p(y_{it}, s_{it} = 1|\alpha_i, \eta_i, \boldsymbol{x}_{it}, \boldsymbol{w}_{it})}{p(s_{it} = 1|\eta_i, \boldsymbol{w}_{it})}.$$

Using the same approximation in Equation (2.6), we can model the conditional probability for observed response variables y_{it} as

$$p^*(y_{it} = 1|s_{it} = 1) = \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + q^1_{it}\rho), \tag{2.15}$$

where

$$q^1_{it} = \mathrm{E}(u_{it}|u_{it} > -\bar{u}_{it}) = -(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma}) - \frac{\ln[1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]}{\Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})}, \tag{2.16}$$

and again, Equation (2.15) describes a fixed-effects logit model with a correction term corresponding to the conditional expected value of the residuals of the selection equation (2.11).

2.4 Pseudo conditional likelihood estimator

With fixed-T panel data, a fixed-effects approach to the estimation of the parameters of the logit model is based on the maximization of the conditional likelihood, (Chamberlain, 1980) given suitable sufficient statistics for the individual intercepts, in order to avoid the incidental parameters problem (Neyman and Scott, 1948). The conditional maximum likelihood estimator, which is common practice for static binary panel data models, can be adapted for the estimation of the model formulation proposed in Section 2.3.

Define now $\boldsymbol{\theta}$ as the vector collecting the set of parameters in the approximating models in Equations (2.7) and (2.15). Then, we have $\boldsymbol{\theta} = (\boldsymbol{\beta}', \delta, \rho)'$ for the endogenous dummy model in Equation (2.7). Similarly, we define

$\boldsymbol{\theta} = (\boldsymbol{\beta}', \rho)'$ for the approximation concerning the sample selection issue in Equation (2.15). We consider the joint distribution of the response configuration $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, conditional on the individual effect α_i, the set of model covariates $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})$, the correction term $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{iT})'$, and the endogenous selection variable $\boldsymbol{s}_i = (s_{i1}, \ldots, s_{iT})'$. It's probability can be written as

$$p^*(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, \boldsymbol{q}_i, \boldsymbol{s}_i) = \frac{\exp\left[y_{i+}\alpha_i + \sum\limits_{t=1}^{T} y_{it}\left(\boldsymbol{z}_{it}'\boldsymbol{\theta}\right)\right]}{\prod\limits_{t=1}^{T}\left[1 + \exp\left(\alpha_i + \boldsymbol{z}_{it}'\boldsymbol{\theta}\right)\right]},$$

where $\boldsymbol{z}_{it} = (\boldsymbol{x}_{it}', s_{it}, q_{it})'$ or $\boldsymbol{z}_{it} = (\boldsymbol{x}_{it}', q_{it})'$ for the cases of self- and sample selection, respectively, and collects the observation of the t-th time period of the covariates. Moreover, $y_{i+} = \sum_{t=1}^{T} y_{it}$ is the *total score*, which is a sufficient statistic for the incidental parameter α_i. Therefore, the joint probability of $\boldsymbol{y}_i$, conditional on $\boldsymbol{X}_i$, $\boldsymbol{q}_i$, $\boldsymbol{s}_i$, and y_{i+}, is

$$p^*\left(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{q}_i, \boldsymbol{s}_i, y_{i+}\right) = \frac{\exp\left[\sum\limits_{t=1}^{T} y_{it}\left(\boldsymbol{z}_{it}'\boldsymbol{\theta}\right)\right]}{\sum\limits_{\boldsymbol{d}:d_+=y_{i+}} \exp\left[\sum\limits_{t=1}^{T} d_t\left(\boldsymbol{z}_{it}'\boldsymbol{\theta}\right)\right]}, \tag{2.17}$$

which no longer depends on α_i and where the sum at the denominator is extended to all the possible response configurations $\boldsymbol{d}$ such that $d_+ = y_{i+}$, where $d_+ = \sum\limits_{t=1}^{T} d_t$.

The formulation of the conditional log-likelihood corresponding to Equation (2.17) relies on the fixed quantities q_{it}, that are based on a preliminary estimation of the parameters in the equation for the endogenous dummy/selection variable s_{it}. The estimation approach is therefore based on two-steps:

Step 1 Estimates of the regression parameters $\boldsymbol{\gamma}$ in the equation for the selection variable are needed to compute the correction term q_{it} in Equation (2.7) or Equation (2.15).

As the error term u_{it} follows a standard logistic distribution, the joint probability for the response configuration of the selection variable $\boldsymbol{s}_i$, conditional on the individual effect η_i and the covariates $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{it})$, is

$$p(\boldsymbol{s}_i|\eta_i, \boldsymbol{W}_i) = \frac{\exp\left[s_{i+}\eta_i + \sum_{t=1}^{T} s_{it}(\boldsymbol{w}'_{it}\boldsymbol{\gamma})\right]}{\prod_{t=1}^{T}\left[1 + \exp\left(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma}\right)\right]},$$

where $s_{i+} = \sum_{t=1}^{T} s_{it}$ is again a *total score* and a sufficient statistic for the incidental parameter η_i. The joint probability conditional on $\boldsymbol{W}_i$ and s_{i+} is therefore

$$p\left(\boldsymbol{s}_i|\boldsymbol{W}_i, s_{i+}\right) = \frac{\exp\left[\sum_{t=1}^{T}(s_{it}\boldsymbol{w}_{it})'\boldsymbol{\gamma}\right]}{\sum_{\boldsymbol{b}:b_+=s_{i+}} \exp\left[\sum_{t=1}^{T} b_t(\boldsymbol{w}'_{it}\boldsymbol{\gamma})\right]},$$

which no longer depends on η_i and where the sum at the denominator is extended to all the possible response configurations $\boldsymbol{b}$ such that $b_+ = s_{i+}$, where $b_+ = \sum_{t=1}^{T} b_t$. Estimation of $\boldsymbol{\gamma}$ is obtained by the following conditional log-likelihood function

$$\ell(\boldsymbol{\gamma}) = \sum_{i=1}^{n} 1\{0 < s_{i+} < T\} \log p\left(\boldsymbol{s}_i|\boldsymbol{W}_i, s_{i+}\right),$$

which can be maximized by a Newton-Raphson algorithm. Given the estimated regression parameters $\hat{\boldsymbol{\gamma}}$, the estimate of η_i is obtained by Maximum Likelihood[5] and the correction term $\hat{w}_{it}$ can be derived by simply substituting $\hat{\gamma}$ and $\hat{\eta}_i$ in Equations (2.9) and (2.10) or Equation (2.16).

Step 2 Let $\hat{\boldsymbol{\psi}}$ collect the estimators obtained in the first step, $(\hat{\gamma}', \hat{\eta}_i)'$. The set of parameters $\boldsymbol{\theta}$ is estimated by maximizing the conditional log-likelihood of (2.17), which can be written as

$$\ell^*(\boldsymbol{\theta}|\hat{\boldsymbol{\psi}}) = \sum_{i} 1\{0 < y_{i+} < T\} \log p^*_{\boldsymbol{\theta}|\hat{\boldsymbol{\psi}}}(\boldsymbol{y}_i|\boldsymbol{X}_i, \hat{\boldsymbol{w}}_i, \boldsymbol{s}_i, y_{i+}).$$

[5] Given the value of $\hat{\boldsymbol{\gamma}}$, estimated intercepts are computed maximizing individual likelihood contributions, $\log p(\boldsymbol{s}_i|\eta_i, \boldsymbol{W}_i)$ with respect to the parameter η_i.

The function $\ell^*(\boldsymbol{\theta}|\hat{\boldsymbol{\psi}})$ may be maximised by a Newton-Raphson algorithm and the resulting $\hat{\boldsymbol{\theta}}$ is the PCML estimator (PCML). This estimator is proved to be consistent and asymptotically normal, as argued by Andersen (1970).

Moreover, this result is still valid in the two-step procedure. Following Newey and McFadden (1994), the second step estimator is not affected by the generated regressor q_{it}. This happens because, under the set of assumptions, the probability limit of the estimator exploited in the first step is non-stochastic and coincides with the true parameters in Equation (2.3) or Equation (2.13), namely $\hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$.

However, the model is derived by an approximation of the joint probability of (s_{it}, y_{it}) around $\rho = 0$ and differs from the true one in presence of correlation. Therefore, the PCML estimator is consistent for a set of pseudo-true parameters $\boldsymbol{\theta}^*$, that equals the true parameters of the model in Equation (2.4) or Equation (2.14), denoted by $\boldsymbol{\theta}_0$, only when $\rho = 0$. We then observe an asymptotic bias, $(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)$, increasing in the true value of the correlation coefficient ρ_0.

2.5 Testing for endogenous self and sample selection

As mentioned in Section 2.3.1 and Section 2.3.2, the correlation between the error terms of the two equations in the model leads to the endogenous selection mechanism. This section shows how to test for endogeneity in both cases in a simple way. Given the estimates of the two approximating models, we can test for the null hypothesis of exogeneity of the dummy/absence of endogenous selection, $H_0 : \rho = 0$, against the the bidirectional hypothesis, $H_0 : \rho \neq 0$. It is worth recalling that the PCML estimator is consistent for the true parameters of the model under the null hypothesis.

The simplest way to test the null hypothesis consists in computing the usual t-test on the coefficient, given by

$$t = \frac{\hat{\rho}}{s.e._{\hat{\rho}}} \overset{a}{\sim} t_{nT-k-1},$$

where $s.e._{\hat{\rho}}$ denotes the standard error of the parameter, k is the dimension of $\boldsymbol{\theta}$ and the t-statistic is asymptotically distributed as a t-distribution with $nT-k-1$ degrees of freedom. In this kind of two-step estimation procedure, it is a common practice to correct standard errors, taking into account the first-step estimates (Newey and McFadden, 1994). In this regard, we rely on a Generalized Method of Moments (GMM) approach similar to the one developed in Bartolucci and Nigro (2012). The estimating equations of the first and the second steps are jointly considered as unique system of the type:

$$g(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}) = \sum_i \mathbb{1}\{0 < s_{i+} < T\}\mathbb{1}\{0 < y_{i+} < T_i\} \begin{pmatrix} \nabla_{\boldsymbol{\pi}} \ell_i(\hat{\boldsymbol{\psi}}) \\ \nabla_{\boldsymbol{\theta}} \ell_i^*(\boldsymbol{\theta}|\hat{\boldsymbol{\psi}}) \end{pmatrix} = \mathbf{0},$$

where ∇_* denotes the derivatives of the likelihood w.r.t. a set of parameters. Starting from this system, a first-step robust estimator for the covariance matrix is obtained by the following expression,

$$W(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) = H(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})^{-1} \left[S(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) \right] H(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})^{-1},$$

where $H(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$ is defined as

$$H(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) = \sum_i \mathbb{1}\{0 < s_{i+} < T\}\mathbb{1}\{0 < y_{i+} < T_i\} \begin{pmatrix} \nabla_{\boldsymbol{\pi}\boldsymbol{\pi}} \ell_i(\hat{\boldsymbol{\psi}}) & \mathbf{0} \\ \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}} \ell_i^*(\boldsymbol{\theta}|\hat{\boldsymbol{\psi}}) & \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_i^*(\boldsymbol{\theta}|\hat{\boldsymbol{\psi}}) \end{pmatrix},$$

where $\mathbf{0}$ is a null square matrix and

$$S(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) = \sum_i g_i(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) g_i(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})'$$

is the outer product of the elements of the vector $g(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}})$. Finally, the standard errors for $\hat{\boldsymbol{\theta}}$ are the elements in the main diagonal of the suitable lower-right sub-matrix of $W(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$.

However, as suggested by Semykina and Wooldridge (2018), there is no need to use first-step robust standard errors, since the the standard estimator of the covariance matrix is consistent under the null hypothesis.

Further to the t-test, the LR test is an attractive alternative, as well as all the standard tests bases on the ML estimator, which are common practice in cross-sectional studies. It is worth noting that in this case it is not possible to compute the LR test by comparing the conditional log-likelihood function of the approximating model and the one of the standard logit model that ignores the selection mechanism. In fact, the subjects that actually contribute to the two log-likelihood functions are different because of the two step estimation strategy shown in Section 2.4. Namely, we exclude from the computation of the likelihood functions all the subjects without variation in the outcome variable, while the PCML also gets rid of individuals without variation in the selection variable. In any case, the LR test can be based on the approximating model and it is given by the difference of the log-likelihood functions of the full model $\ell^*(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\psi}})$, and the restricted one $\ell^*(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\psi}}, \rho = 0)$. The LR statistic is defined as

$$LR = -2[\ell^*(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\psi}}, \rho = 0) - \ell^*(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\psi}})] \overset{a}{\sim} \chi^2_1, \tag{2.18}$$

and it is asymptotically distributed as a chi-squared with one degree of freedom (Andersen, 1971).

In order to compare the proposed test with alternative provided by existing literature, we describe a simple alternative two-step testing procedure that relies on the CRE approach. For the case of binary endogenous dummy, the test is based on a two-step procedure described in Lin and Wooldridge (2017) where the unobserved heterogeneity is dealt with the Mundlak's device (Lin and Wooldridge, 2019). The two steps are:

1. Estimate a pooled probit for s_{it}, augmented with the individual time averages of exogenous variables, $\bar{\boldsymbol{x}}_i$, and a set of time dummies. Then, compute the inverse Mill's ratio.

2. Use a pooled probit for y_{it} where the Mill's ratio, related to parameter ρ, and $\bar{\boldsymbol{x}}_i$ are additional regressors .

The null hypothesis $H_0 : \rho = 0$ can be tested by a simple t-test on the estimated coefficient. The test for the sample selection is very similar and has been proposed by Semykina and Wooldridge (2018). The difference with the estimation shown above is that the second step estimation is restricted to the observations for which $s_{it} = 1$.

2.6 Dynamic Logit Model

The formulations in the previous sections concern static logit models. However, the extension of the results to dynamic models is straightforward. We include the lagged dependent variable, $y_{i,t-1}$, in the Equation (2.4) and we get

$$y^*_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + s_{it}\delta + y_{i,t-1}\phi + \varepsilon_{it}$$

where ϕ is the state dependence parameter.

Clearly, the identification imposes to lose the first observation of individual outcomes, denoted y_{i1}, which enters the conditioning set and is treated as given. Following the same approximation adopted for the static model will lead us, for the case of endogenous explanatory variable, to an approximated conditional probability given by

$$p^*(y_{it}|s_{it} = 1, \alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}, y_{i1}) = \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + s_{it}\delta + \phi y_{i,t-1} + q_{it}\rho),$$

that is a DL model, where all the other variables and parameters are defined in Section 2.3.1.

This methodology is also applied to the case of sample selection, even though we have to take into account a peculiarity. In fact, since the outcome y_{it} is observed only when $s_{it} = 1$, it could happen that, even if y_{it} is observed, $y_{i,t-1}$ is not available. Hence, the state dependence parameter is not properly identified.

A feasible way to overcome this problem is to consider, for each $i-$th subject in the sample, all the possible sequences of at least three consecutive not-missing observations. This sequences will autonomously contribute to the likelihood function, regardless the original subject they belong. The first element of each group of consecutive observations will be treated as given, and it will be the initial condition. The second point is about the individual intercepts. When two subgroups come from the same subjects it is not clear how to manage the subgroups common parameter α_i. However, there is no need to formulate hypotheses about this aspect because of the conditional estimation. Therefore, the identification strategy for the sample selection model leads to

$$p^*(y_{it}|s_{it}=1,\alpha_i,\boldsymbol{x}_{it},y_{i,t-1},y_{i1}) = \Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\phi y_{i,t-1}+q_{it}\rho).$$

Finally, the estimation procedure relative to both self and sample selection models is slightly different from the one presented in Section 2.4. Differently from the static logit model, the DL does not admit a sufficient statistic for the incidental parameters α_i. Therefore, we adopt the estimator proposed by Bartolucci and Nigro (2012) based on a QE model that acts as an approximation and that admits the total score as a sufficient statistic for the incidental parameters.

2.7 Simulation study

This section illustrates the design and discuss the main results of the simulation study used to investigate the power properties of the tests described in Section 2.5.

2.7.1 Endogenous Binary Explanatory Variable

The simulation study is based on samples subject to the observational rule in Section 2.3.1. For the pairs of binary data (s_{it}, y_{it}), the model assumes that

$$s_{it} = \mathbb{1}\{\eta_i+\gamma_0+x_{it}\gamma_1+m_{it}\gamma_2+u_{it}>0\}, \tag{2.19}$$
$$y_{it} = \mathbb{1}\{\alpha_i+x_{it}\beta+s_{it}\delta+\varepsilon_{it}>0\}. \tag{2.20}$$

The covariates x_{it} and m_{it} are normally distributed with zero mean and variance equal to $\pi^2/3$. The error terms ε_{it}, defined as in Equation (2.5), are drawn from a Gaussian copula, where u_{it} and v_{it} follow a logistic distribution with zero mean and variance equal to $\pi^2/3$. In order to evaluate the performance of the test for dynamic models, we build the latter simply by adding $y_{i,t-1}\phi$ in the linear index in Equation (2.23), where $y_{i,t-1}$ is the lagged value of the response variable and ϕ is the state dependence parameter. Moreover, we define the initial observation as in Equation (2.22). Then, the data generating process for the dynamic model is given by

$$s_{it} = \mathbb{1}\{\eta_i + x_{it}\gamma_1 + m_{it}\gamma_2 + u_{it} > 0\}, \tag{2.21}$$

$$y_{i1} = \mathbb{1}\{\alpha_i + x_{i1}\beta + s_{i1}\delta + \varepsilon_{i1} > 0\}, \tag{2.22}$$

$$y_{it} = \mathbb{1}\{\alpha_i + x_{it}\beta + s_{it}\delta + y_{i,t-1}\phi + \varepsilon_{it} > 0\} \quad \text{for} \quad t \geq 2. \tag{2.23}$$

For the individual intercepts we define three different designs. All the scenarios allow for correlation between α_i, η_i, and the covariates x_{it}. The benchmark design fits the assumptions of Mundlak (1978), and both individual effects are given by the sum of the time averages of x_{it} and m_{it} for each subject, plus an additional idiosyncratic term, ξ_i, drawn form a gaussian distribution, that is

$$\alpha_i = \frac{1}{T}\sum_{t=1}^{T} x_{it} + \frac{1}{T}\sum_{t=1}^{T} m_{it} + \xi_{1i}, \quad \eta_i = \frac{1}{T}\sum_{t=1}^{T} x_{it} + \frac{1}{T}\sum_{t=1}^{T} m_{it} + \xi_{2i}.$$

In the second and the third cases, along the lines of Honoré and Kyriazidou (2000), α_i is given by the average of the first three observations x_{it} for each configuration and η_i is drawn from a Gaussian copula, so that the two designs are

$$\alpha_i = \frac{1}{3}\sum_{t=1}^{3} x_{it}, \quad \eta_i = 0.5\,\alpha_i + \sqrt{0.75}\,\xi_i, \tag{2.24}$$

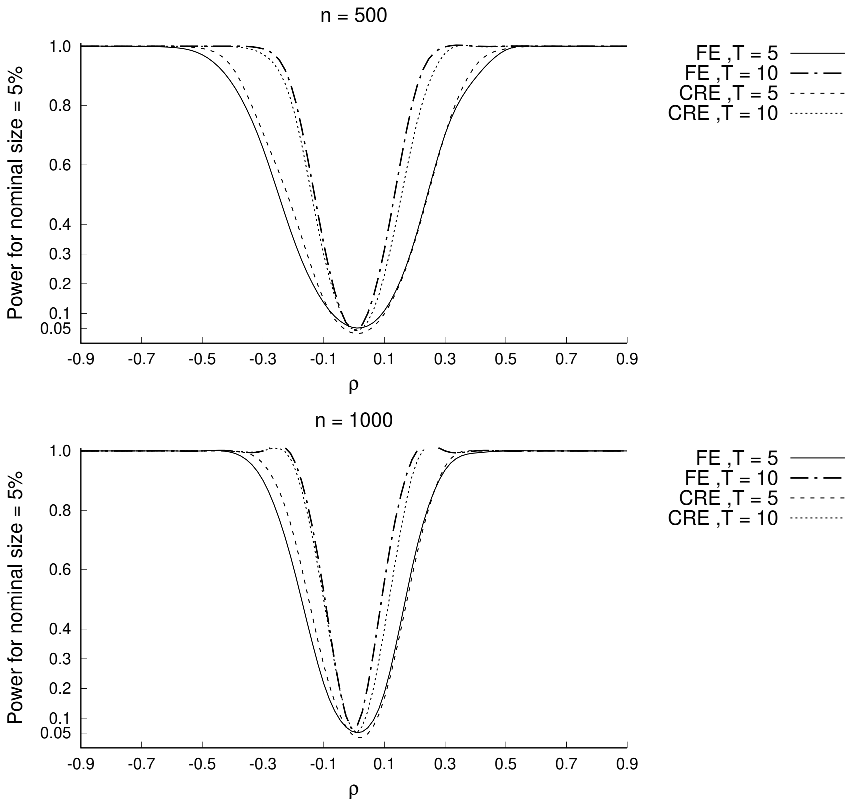
and

$$\alpha_i = \frac{1}{3}\sum_{t=1}^{3} x_{it}, \quad \eta_i = 0.8\,\alpha_i + \sqrt{0.36}\,\xi_i, \tag{2.25}$$

where, again, $\xi_i \sim N(0,1)$. These two designs for unobservables in Equations (2.24) and (2.25) will be referred as HK1 and HK2, respectively. The only difference between HK1 and HK2 is the correlation between the unobservables which indirectly defines the correlation between x_{it} and η_i.

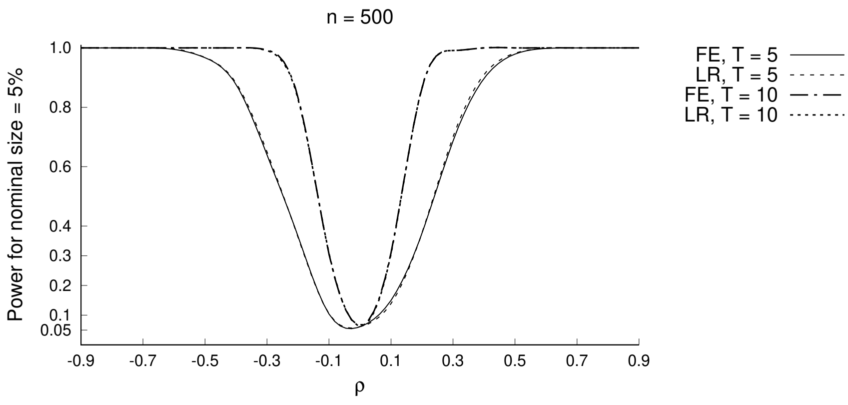
Throughout the simulation study, we set $\beta = \delta = 1$ in Equations (2.20) and (2.23), as well as γ_1 and γ_2 in Equation (2.19) and (2.21). The parameter γ_0 in Equation (2.19) allows us to control the proportion of observations of the selection variable that are equal to 1. As a benchmark, we set $\gamma_0 = 0$ so that this proportion is close to the 50%. Alternatively, we consider the case of $\gamma_0 = 5$ that produces a 90% of $s_{it} = 1$. The state dependence parameter ϕ, assumes values in $\{0.5, 1\}$. Finally, we let ρ vary from -0.9 to 0.9, by steps of 0.1. The sample sizes considered are $n = 500, 1000$ for $T = 5, 10$ and the number of Monte Carlo replications is 1000.

Figure 2.1: Endogenous Dummy: t-test, ($H_1 : \rho \neq 0$), Benchmark Design



We now consider the power properties of the tests described in Section 2.5, for the null hypothesis of non-endogeneity of s_{it}, namely $H_0 : \rho = 0$. The

Figure 2.2: Endogenous Dummy: LR-test, ($H_1 : \rho \neq 0$), Benchmark Design

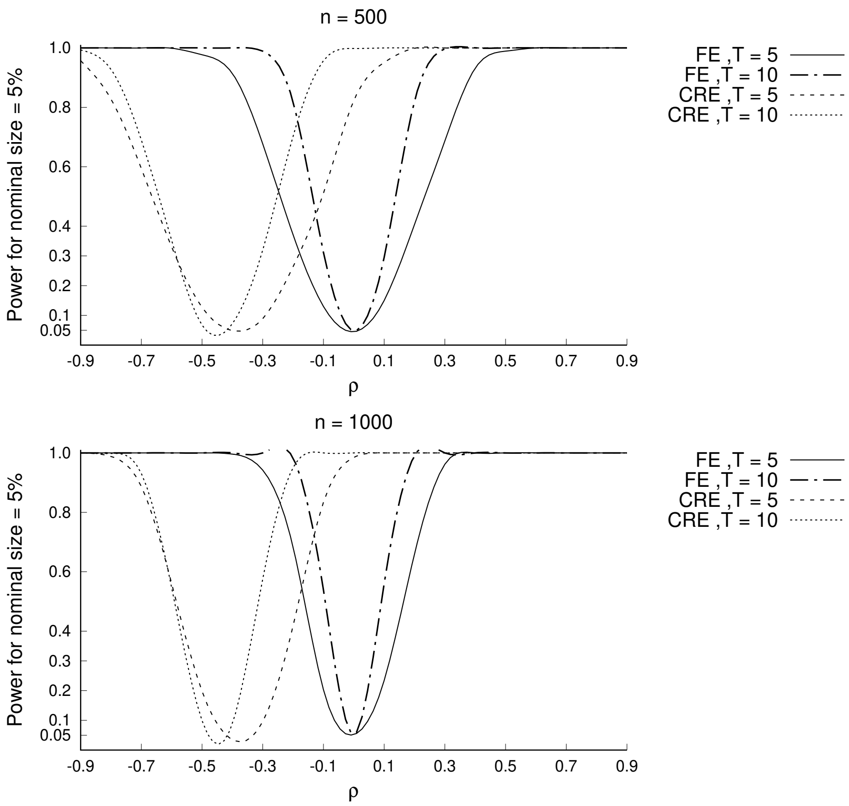


nominal size of the test is of 0.05 and the curves are obtained by cubic spline interpolation. The following figures report the results for the t-test denoted FE, and for the LR test (LR) based on the approximating model proposed in this work. Further, we report the results for t-test based on estimates that exploit the CRE approach. It is worth recalling that the benchmark design is built such that the Mundlak device works perfectly.

Figure 2.1 shows the rejection rate of the t-tests for the benchmark design. Under this assumption, the two approaches should be equivalent. First of all, the tests always attain the nominal size in all the sample configurations. Moreover, the power curves for the two approaches are very close. Obviously, the larger the sample in n or T, the higher the rejection rate of the tests when the alternative hypothesis is true. Figure 2.2 compares the results of the FE t-test and the LR test for $n = 500$. As we can see, the tests exhibit equivalent results since the size and the average rejection rates are the same.

As mentioned in Section 2.3, one of the main advantages of the approach presented in this work is the absence of assumptions about the structure of the individual effects. In this regard, Figure 2.3 includes the power curves of the tests when data are generated from the HK1 design. We observe that the proposed methodology is robust to the different data generating process. Again, it attains the size and the test gains power as the number

Figure 2.3: Endogenous Dummy: t-test, ($H_1 : \rho \neq 0$), HK1 Design



of observations increases. On the contrary, the CRE approach fails to fully model the correlation between the unobservables and the regressors because the Mundlak's device is misspecified by construction. This fact breaks down the ability of the test to detect the correlation between ε_{it} and u_{it}, so that the resulting power curve is shifted due to the adopted design. In this regard, Figure 2.4 reports the results for the HK2 design. In this case the shift of the power curves for the CRE t-test is more severe because unobservables are more correlated than in HK1 design.

A further aspect that is worth mentioning is the effect on the test of large proportions of 1 or 0 in the selection variable. Figure 2.5 includes the power curves of the FE t-test with a proportion of 1 in the selection variable close to 90%. We observe that the test exhibits less power with respect to

Figure 2.4: Endogenous Dummy: t-test, $(H_1 : \rho \neq 0)$, HK2 Design

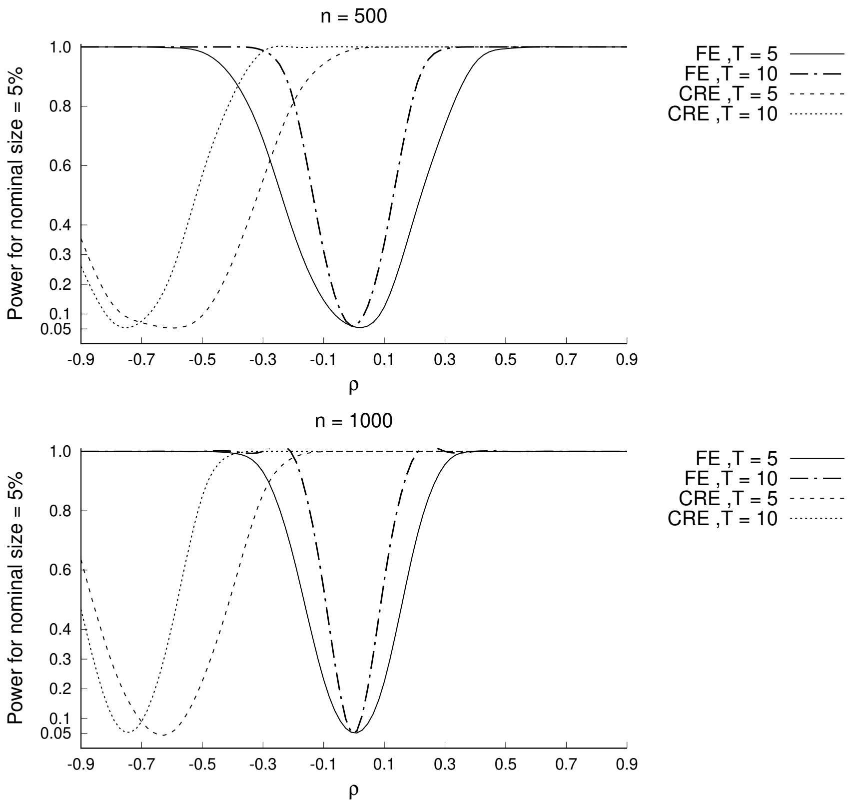


Figure 2.5: Endogenous Dummy: t-test, $(H_1 : \rho \neq 0)$, 90% of $s_{it} = 1$

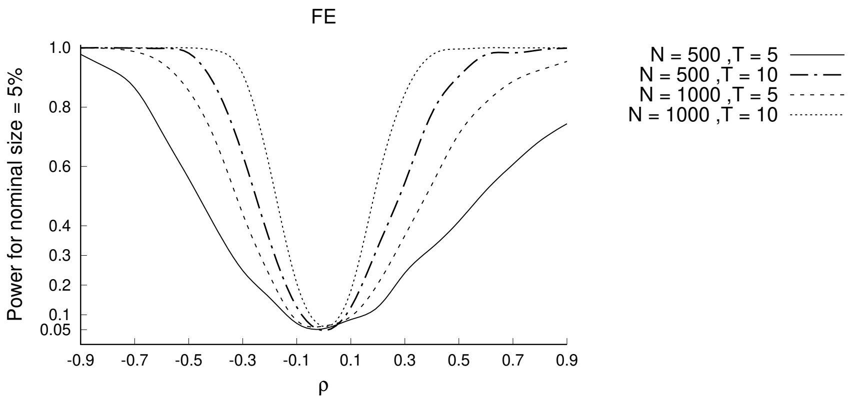
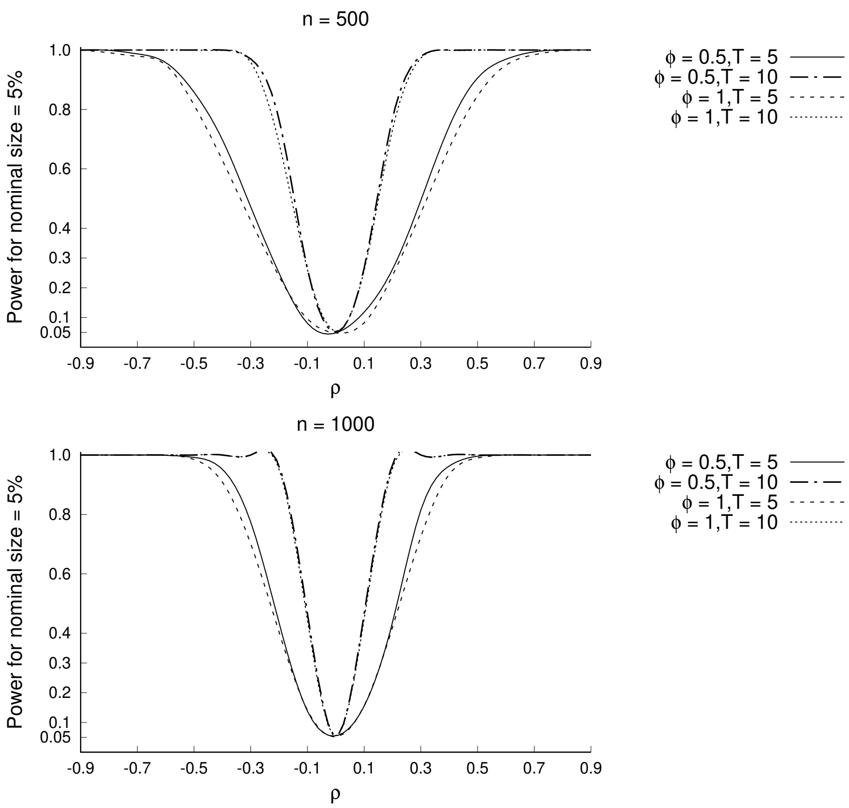


Figure 2.6: Endogenous Dummy: t-test, $(H_1 : \rho \neq 0)$, Dynamic Logit



the benchmark design in Figure 2.1 because in this case the actual sample size is smaller. This fact is due to the first step estimation. Considerable imbalances in the proportion of 1 and 0 in s_{it} imply low variability of the selection variable. Consequently, the conditional estimation in the first step require to exclude a large number of observation that are not informative.

Finally, Figure 2.6 reports the results about the t-test based on $\hat{\rho}$ for the DL model with the two different values for the state dependence parameter. Even though parameters are estimated by an approximating model (Bartolucci and Nigro, 2012) the testing procedure is still valid. Also in this case the test attains the nominal size under the null hypothesis H_0. However, we can observe a loss of power in the case of $T = 5$ with respect to the static model, while the tests produce similar results for $T = 10$. A peculiar aspect is given comparing the curves for different levels of the state dependence parameter. Despite the estimator proposed in Bartolucci and Nigro (2012) is proved to consistent only for the case $\phi = 0$, the power curves are very close for the two different level of state dependence.

2.7.2 Sample Selection

The simulation design adopted to study the problem of sample selection is close to the one shown in the previous section. It is based on samples subject to the observational rule in Equations (2.11) and (2.12). The outcome variables, (s_{it}, y_{it}), are modeled by

$$\begin{aligned} s_{it} &= \mathbb{1}\{\eta_i + \gamma_0 + x_{it}\gamma_1 + z_{it}\gamma_2 + u_{it} > 0\}, \qquad (2.26)\\ y_{it} &= \mathbb{1}\{\alpha_i + x_{it}\beta + \varepsilon_{it} > 0\}, \end{aligned}$$

where y_{it} is observed only if $s_{it} = 1$. The covariates, the individual intercepts and the parameters are the same shown in Section 2.7.1. Also in this case, dynamic models include the lagged response variable and the state dependence parameter and the initial condition,

$$\begin{aligned} s_{it} &= \mathbb{1}\{\eta_i + \gamma_0 + x_{it}\gamma_1 + z_{it}\gamma_2 + u_{it} > 0\}, \\ y_{i1} &= \mathbb{1}\{\alpha_i + x_{i1}\beta + \varepsilon_{i1} > 0\}, \\ y_{it} &= \mathbb{1}\{\alpha_i + x_{it}\beta + y_{i,t-1}\phi + \varepsilon_{it} > 0\} \quad \text{for} \quad t \geq 2. \end{aligned} \tag{2.27}$$

In this case, we have different values for the intercept γ_0 in Equations (2.26) and (2.27), it assumes values in $\{-1, 1, 3\}$, providing about 60%, 40%, and 20% censored observations in y_{it}, respectively. The sample sizes considered are $n = 500, 1000$ for $T = 5, 10$ for static models. Since the identification strategy described in Section 2.6 requires, for a unit in the sample, at least three consecutive observations and reduces the actual sample size considered, we explore only samples with $n = 500, 1000$ and $T = 10$ for dynamic models.

Sample selection implies unbalanced panel datasets. Hereafter we will consider the benchmark design for the unobservable components. A fundamental aspect we take into account is the number of missing observations. At first, we consider a censoring level of 40%. Figure 2.7 shows the usual t-tests, the one proposed in Section 2.5 (FE) and the Semykina and Wooldridge (2018) proposal (CRE). Also for the sample selection case, the test behaves as expected in terms of size and power. The main difference with respect to case of the endogenous dummy is the actual sample size exploited in the estimation procedure. In fact, the curves tend to be less steep due to the smaller number of available observations.

Figure 2.8 reports the results for the FE test and the alternative LR test with $n = 500$. As well as in the case of the endogenous dummy, the power curves of the test almost perfectly overlap. Further, we observe that the performance of the test is affected by the level of censoring and this effect is clear in Figure 2.9, where we can find the rejection rate of the proposed FE t-test for two different levels of missing observations: 20% and 60%, where the test exhibits different power curves. As well as the case of endogenous dummy described in Figure 2.5, the lower variability of the selection variable, the smaller the number of available observations in the second step. Despite the first case implies a larger number of available observations of y_{it} with

Figure 2.7: Sample Selection: t-test, $(H_1 : \rho \neq 0)$, Benchmark Design

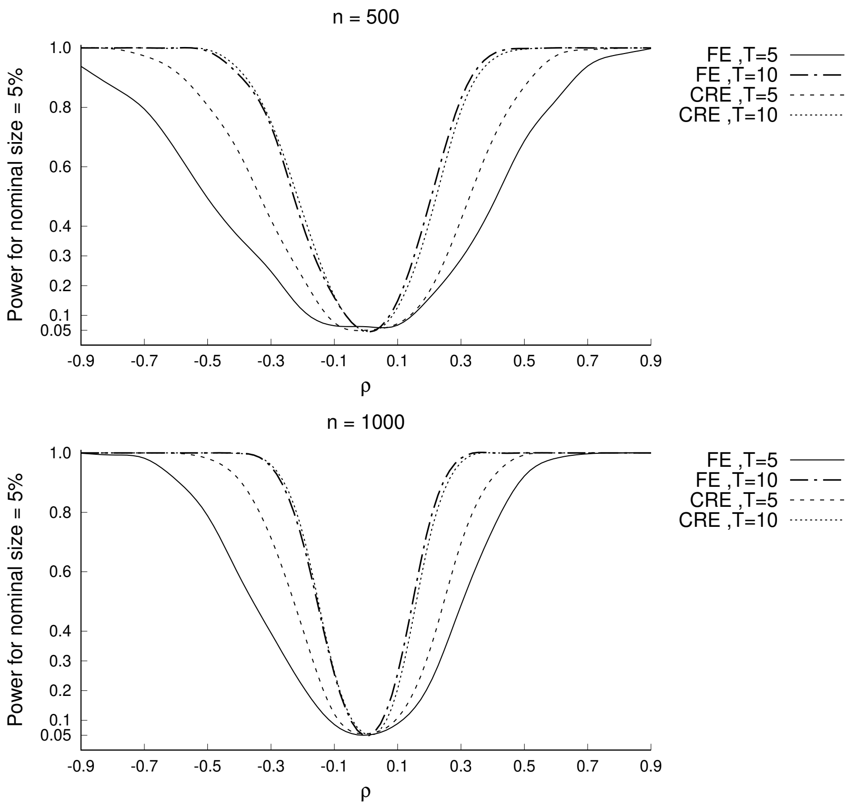


Figure 2.8: Sample Selection: LR-test, $(H_1 : \rho \neq 0)$, Benchmark Design

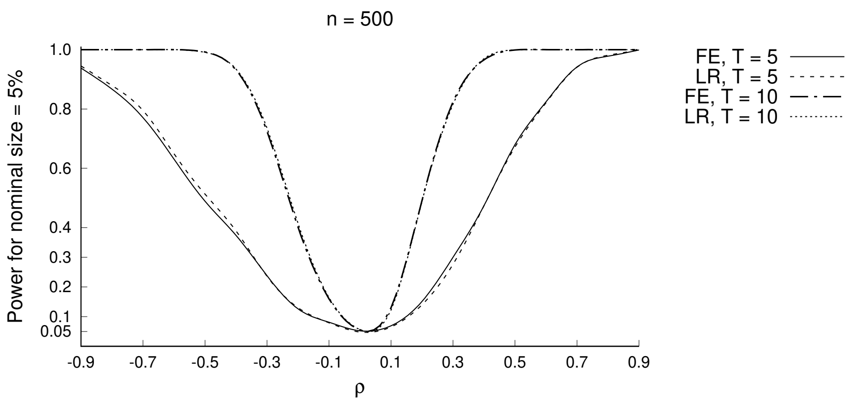
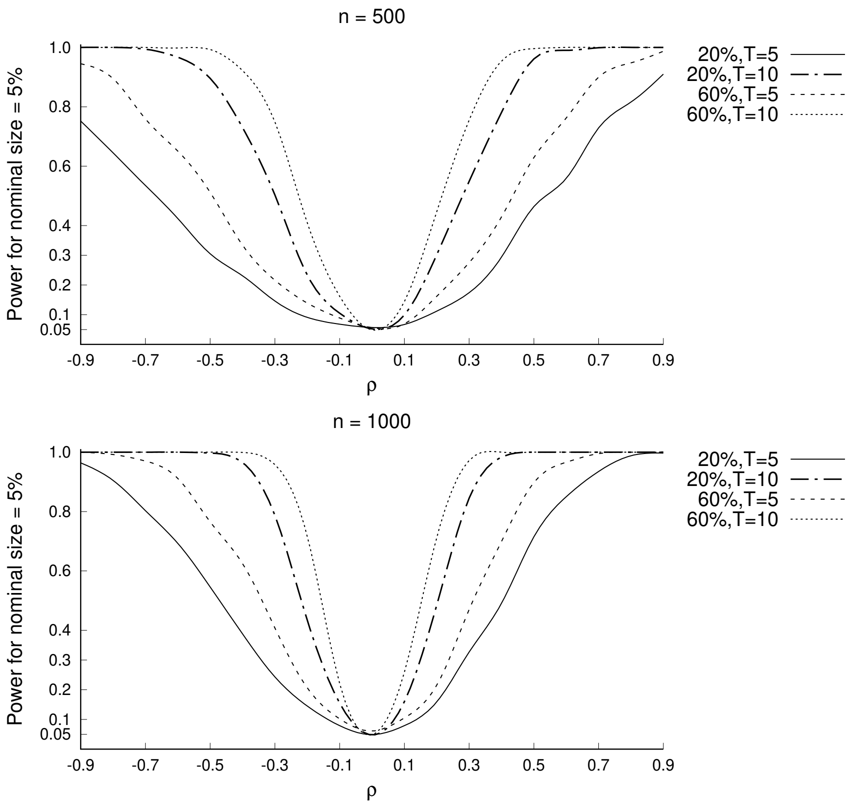


Figure 2.9: Sample Selection: t-test, $(H_1 : \rho \neq 0)$, Alternative censoring levels



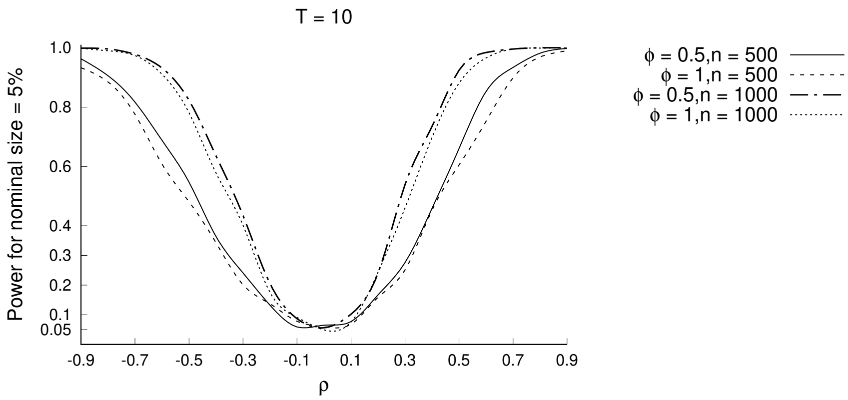
respect to the second scenario, the actual sample size available for estimation procedure is larger in the "60%−"scenario.

Finally, the results for the dynamic setup are satisfying in terms of size and power. Figure 2.10 reports the curves for the t-test (FE), with a 40% of missing observations. Clearly, the test has a smaller rejection rate compared with the static model because of the identification strategy.

2.8 Application

This section illustrates an application of the proposed methodology to real data and focuses on the analysis concerning the impact of retirement on

Figure 2.10: Sample selection: t-test, $(H_1 : \rho \neq 0)$, Dynamic Logit



the health conditions. The empirical illustration exploits the Survey of Health, Ageing and Retirement in Europe (SHARE). This dataset provides individual-level information about health, socio-economic status and social and family networks for a set of individuals in 27 European countries and Israel. The actual sample here considered is referred to a balanced panel of 8,753 individuals with complete questionnaires observed in Waves 1,2,4 and 5 of the dataset, for a total of 35,012 observations.[6]

The SHARE dataset has become very popular in the recent literature and it has been employed in different studies that empirically investigate the relation between retirement and health. Among others, Coe and Zamarro (2011) exploit different measures concerning individual health status and find that “retirement has a health preserving effect on the overall general health”. Conversely, Mazzonna and Peracchi (2012, 2017) show the negative impact of retiring on health conditions and cognitive abilities together with a strong heterogeneity of these effects “across occupational groups” and between the short- and the long-run. Moreover, other studies investigated the impact of retirement on financial hardship (Angelini et al., 2009), unhealthy behaviours

[6]The sample is restricted to individuals interviewed in Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Switzerland and Belgium. Excluding Greece, whose data are not available for the fourth and the fifth waves, these countries are the same considered by Coe and Zamarro (2011).

(Celidoni et al., 2017), and social relationships (Comi et al., 2019).

A common issue in these studies concerns the potential endogeneity of the retirement status. Eibich (2015) argued about three different sources of endogeneity. First of all there could be reverse causality. Since health conditions affects the retirement decisions, one might expect that individuals with bad health conditions have a higher probability to retire early. Secondly, there could be a "justification bias" meaning that people who retired early could underreport their health status in order to justify the fact they are no longer in the labour force. In a study of the impact of the retirement on the probability of being in bad health conditions, these phenomena could lead to an upward-bias of the resulting coefficient. Third, an omitted variable bias due to time-constant unobservables has to be taken into account.

In this application, the main outcome BH_{it} is a binary variable that is equal to one when an individual is in *bad health* conditions. This variable is generated starting from a self-assessed health status in scale that ranges in *Excellent*, *Very Good*, *Good*, *Fair*, and *Poor*.[7] The dichotomization of the self reported health conditions is popular in the literature and provides an easy interpretation of the sign of the marginal effects (Eibich, 2015). We will adopt two different definitions of bad health, the first variable $BH1_{it}$ equals one when the reported health status is *Fair* or *Poor* and zero otherwise. This definition of bad health status a standard practice in the literature (Coe and Zamarro, 2011). In the second version, $BH2_{it}$ is one also when the self-assessed condition is *Good*. In order to perform our analysis, we take into account the following two-equation model

$$\begin{aligned} & BHj_{it} \quad j=1,2 \quad i=1,...,8753 \quad t=1,...,4 \\ & BHj_{it} = \mathbb{1}\{\alpha_i + \beta_1 Ret_{it} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} > 0\}, \\ & Ret_{it} = \mathbb{1}\{\eta_i + \gamma_1 EligE_{it} + \gamma_2 EligN_{it} + \boldsymbol{x}_{it}'\boldsymbol{\gamma} + u_{it} > 0\}, \end{aligned} \tag{2.28}$$

where we assume that the binary explanatory variable Ret_{it} represents the retirement status and it is treated as endogenous. This regressor is equal

[7] As pointed out by Coe and Zamarro (2011) this scale is comparable with the U.S. Health and Retirement Study.

to 1 when an individual is actually retired and zero otherwise.[8] The vector $\boldsymbol{x}_{it}$ include a set of control variables: age, age squared, marital status, gender, years of education, number of children, household income, and a set of country-specific and time dummies. Clearly, all time-invariant covariates will be removed from the specification with fixed-effects in order to achieve identification. Hence, the set of explanatory variables is given by age, age squared, household income, and the time dummies for models with fixed-effects.[9]

For what concerns Equation (2.28), the one for the endogenous variable Ret_{it}, we include two over-identifying restrictions, $EligE$ and $EligN$ which are two dichotomous variables that equals one when an individual is eligible for the early or the normal retirement, respectively. Eligibility rules vary across countries and over time, they should have a large explanatory power for the retirement status but at the same time it does not directly affect the individual health status. Finally, we include the presence of unobserved heterogeneity, α_i and η_i which will be treated in different ways according the proposed model specification.

Table 2.1 and Table 2.2 report estimated coefficients for three different specifications of single equation model and their relative two-equation counterpart where we take into account the endogeneity of the retirement, in order to compare the proposed methodology with viable alternatives.

The first three columns report the estimates of a probit model with CRE, a conditional logit model, and a linear probability model with fixed-effects. The columns relative to the testing models include the testing procedures described in Section 2.5 and a fixed-effects linear probability model where the endogeneity of retirement is dealt with a control function approach.[10] In Table 2.3 we can find the estimated coefficients of the identifying restrictions in the equation of retirement for the three different model specification proposed.

[8] Alternative definitions of retirement consider the labour force participation.

[9] The marital status is actually time-varying. Anyway, in order to avoid weak identification due to limited variability over time it is excluded from the FE models. Moreover, the inclusion of this variable does not affect the results.

[10] The two-step least squares is reported here since it is a very popular technique in the literature.

Table 2.1: Estimates Results (BH1)

	Single-Equation Models			Testing Models		
	CRE-Probit	FE-Logit	FE-LPM	CRE	FE	FE-LPM
Bad Health 1						
Retired	-0.062 **	-0.171 ***	-0.022 ***	-0.399 ***	-0.482 **	-0.079 ***
	[0.032]	[0.065]	[0.007]	[0.069]	[0.268]	[0.021]
Age	-0.160 ***	-0.197 ***	-0.035 ***	-0.065 **	-0.122	-0.028 ***
	[0.033]	[0.056]	[0.007]	[0.028]	[0.109]	[0.007]
Age^2	0.001 ***	0.002 ***	0.000 ***	0.000 ***	0.000	0.000 ***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]
Income	-0.000	-0.001 *	-0.000 *	-0.000 ***	0.000	-0.000 *
	[0.000]	[0.000]	[-0.000]	[0.000]	[0.001]	[0.000]
#Children	0.029 **	-	-	0.012 **	-	-
	[0.012]	-	-	[0.005]	-	-
Education	-0.046 ***	-	-	-0.042 ***	-	-
	[0.004]	-	-	[0.002]	-	-
Correction	-	-	-	0.236 ***	0.131	0.065 ***
	-	-	-	[0.040]	[0.094]	[0.022]
t-test on $\hat{\rho}$	-	-	-	5.815	1.223	2.940
(p-value)	-	-	-	(0.000)	(0.164)	(0.000)

[1] CRE models include the full set of explanatory variables to control possible correlation between covariates and unobservables. [2] The FE models only include time-variyng regressors. [3] p-value: *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1

From Table 2.1, all models provide results in line[11] with Coe and Zamarro (2011), emphasizing the negative impact of the retirement on being in bad health and, as expected, this impact is largely underestimated, in absolute value, when we do not take into account endogeneity of the retirement.

Anyway, it is worth mentioning that the test on $\hat{\rho}$ based on the proposed methodology leads to a completely different conclusion with respect to the alternatives, that is we fail to reject the null hypothesis of idiosyncratic endogeneity. This can be due to the different specification of unobservables. It is possible that the CRE and the FE-LPM testing models fail to correctly separate the idiosyncratic and the heterogeneity endogeneity, wrongly detecting as significant the estimated coefficient relative to the "residuals" of the "first stage" regression. Another explanation can be due to the identification strategy of the PCML here proposed. Namely, the actual individuals providing information are those who actually retire and at the same time "switch"

[11]Notice that it is not straightforward to compare these results with those of previous studies since the combinations of these techniques and this sample has not been exploited in the literature.

Table 2.2: Estimates Results (BH2)

	Single-Equation Models			Testing Models		
	CRE-Probit	FE-Logit	FE-LPM	CRE	FE	FE-LPM
Bad Health 2						
Retired	0.019	-0.040	-0.004	-0.308 ***	-0.696 ***	-0.082 ***
	[0.030]	[0.063]	[0.007]	[0.067]	[0.239]	[0.022]
Age	-0.098 ***	-0.118 **	-0.003	-0.033	-0.075	0.007
	[0.031]	[0.055]	[0.007]	[0.026]	[0.127]	[0.007]
Age^2	0.001 ***	0.001 ***	0.000	0.000 *	0.001	-0.000
	[0.000]	[0.065]	[0.000]	[0.000]	[0.001]	[0.000]
Income	-0.000	-0.001 **	-0.000 *	-0.000	-0.000	-0.000 *
	[0.000]	[0.065]	[-0.000]	[0.000]	[0.001]	[0.000]
#Children	0.003	-	-	-0.004	-	-
	[0.012]	-	-	[0.006]	-	-
Education	-0.044 ***	-	-	-0.039 ***	-	-
	[0.004]	-	-	[0.002]	-	-
Correction	-	-	-	0.221 ***	0.245 ***	0.088 ***
	-	-	-	[0.039]	[0.087]	[0.023]
t-test on $\hat{\rho}$	-	-	-	5.639	2.837	3.790
(p-value)	-	-	-	(0.000)	(0.005)	(0.000)

[1] CRE models include the full set of explanatory variables to control possible correlation between covariates and unobservables. [2] The FE models only include time-variyng regressors. [3] p-value: *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1

their health status at least once in the observed sample. It is possible that the main outcome for the subject who retired in the observed period has not sufficiently time variation and that would lead to a weak identification problem.

In order to understand how the definition of the main outcome may affect the estimation procedure, Table 2.2 reports the same set of estimates where the main dependent variable is $BH2_{it}$. Again, even though the interpretation of the parameters is slightly different because of the different definition of the outcome, all results are coherent in underlining the negative impact of retirement on being in bad health. Moreover, all the testing models provide the same results. This seems coherent with the fact that a higher variability of the outcome variable is required for a proper identification. Finally, it is worth mentioning that the coefficients provided by the proposed methodology are not full reliable but represent an approximation of the true set of parameters since the correlation coefficient is different from zero.

Table 2.3: Retired Equation - Excluded Instruments

	CRE	FE	FE-LPM
EligE	0.516 ***	1.000 ***	0.226 ***
	[0.039]	[0.116]	[0.009]
EligN	0.737 ***	1.740 ***	0.271 ***
	[0.035]	[0.114]	[0.009]

2.9 Conclusions

This work introduces a novel method to test for endogeneity of a binary explanatory variable or for endogenous sample selection in binary choice panel data models. Existing methodologies rely on strong parametric assumptions to deal with unobserved heterogeneity. On the contrary, the proposed testing procedure is built in a pure fixed-effects approach, so that unobservables are not assumed to follow a specific distribution.

Moreover, the testing procedure is very easy to implement. It is based on an approximating model that follows a fixed-effects logit model formulation that admits a sufficient statistic for nuisance parameters and allows us to overcome the incidental parameters problem. It is sufficient to perform the conditional maximum likelihood estimation, which is a standard practice in this field, and run a simple variable addition test.

The Monte Carlo study shows that a simple t-test is able to detect endogeneity for both static and dynamic models. It always attains the size under the null hypothesis. In terms of power, it behaves differently according to to the issue we're investigating even tough, in general, it gains power as the endogenous selection becomes more relevant. Further, the test is robust to different structures of individual heterogeneity. The absence of parametric assumptions on unobservables let the test to work well regardless the simulation design. Unfortunately, the alternative testing procedures based on correlated random-effects fail to recognise endogenous selection when unobserved effects are misspecified. Finally, we perform an application to real data. We try to asses the impact of retirement on the health status. The whole set of estimates indicates a negative effect of retirement on the probability of being in bad health conditions even though this effect is underestimated when we consider the retirement as exogenous. The endogeneity of retirement may be due to both reverse causality and unobservables. The proposed methodology allows to take into account both aspects.

Results highlight the advantages and the limits of this approach. The absence of parametric assumptions lets the test to be fully reliable with respect to the role of unobserved heterogeneity but at the same time the main limit of the fixed-effects approach is that data must have enough time variability to achieve identification. Moreover, it is not possible to consistently estimate marginal effects within the conditional inference framework and a recent proposal that aims to fill this gap is the one of Bartolucci and Pigini (2019). Finally, it is worth mentioning that the proposed model cannot be exploited for parameters point estimates when the null hypothesis of the test is rejected, since the PCML here proposed is a bias estimator. Further research should focus on understanding the way approximation bias affects the estimator and on deriving marginal effects for the proposed model.

Chapter 3

Recursive computation of Quadratic Exponential conditional likelihood function

3.1 Introduction

Fixed-effects estimation of non-linear binary choice panel data models plays an important role in the recent econometric literature. A wide set of techniques concerning estimation of the aforementioned models are available. The key idea is to overcome the well-known incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000), discussed in Section 1.2.1, which leads to a bias of the Maximum Likelihood estimator (ML). Consider a longitudinal panel dataset, where we have n individuals observed over a small number T of time occasions. Intuitively, this bias of the ML estimator arises in fixed-effects models because the estimation of the individual specific intercepts is based on the few observations per subject, so that it is not possible to consistently estimate each individual parameter. Moreover, the bias spreads to the estimator of the regression parameters as well, since they are not informationally orthogonal.

A we have seen in Chapter 1, a first approach focuses on the bias correction of the ML estimator (Hahn and Newey, 2004; Fernández-Val, 2009;

Hahn and Kuersteiner, 2011; Dhaene and Jochmans, 2015), corrected likelihood functions (Bester and Hansen, 2009; Arellano and Hahn, 2016; Bartolucci et al., 2016), or corrected score functions (Carro, 2007). The idea of these approaches is to reduce the order of the bias of the ML estimator from $O(T^{-1})$ to $O(T^{-2})$. This class of estimators can be adapted to both probit and logit models in static and dynamic specifications so that these techniques are potentially appealing for a wide set of economic applications. However, bias-corrected estimators show poor finite-sample performance with panel datasets characterised by a small number of time occasions because bias corrections become more precise as T grows. As a rule of thumb, it is necessary that the proportion between the number of subjects and the time occasions they are observed is $n^{1/3} < T$. Moreover, bias corrections are hampered in unbalanced panel datasets. In fact, the subjects with few observations over time make the bias of the MLE and its corrections more severe.

Conditional inference[1] represents an alternative approach. It is based on the existence of sufficient statistics for the incidental parameters. As opposed to the bias-corrected estimators, conditional estimators are model specific but are proved to be fixed-T consistent since these models get rid of the nuisance parameters conditioning the individual likelihood function to the sufficient statistic. Relevant contributions in this field are the one of Chamberlain (1980) for the static logit model. Further, Honoré and Kyriazidou (2000) and Bartolucci and Nigro (2010, 2012) extend the conditional approach to dynamic models such as the Dynamic Logit model (DL) and the Quadratic Exponential (QE) model (Cox, 1972a; Bartolucci and Pennoni, 2007), respectively. From a practitioner perspective, these estimators are appealing for economic applications not only with small-T panel data but also when we have unbalanced data sets. However, conditional estimators require the maximisation of peculiar likelihood functions, whose computational burden limits the applicability of these techniques when T becomes large, so that the parameters estimation is no longer feasible when the likelihood function is computed by standard algebra operations. One way to overcome the computational issues is to exploit recursive algorithms.

[1]Section 1.2.1 for a detailed description.

Many previous works are related to the recursions for a variety of models in the conditional inference framework. A fundamental contribution is by Howard (1972), who shows that the conditional likelihood function of the model proposed in Cox (1972b) is a symmetric function and it can be written recursively. A popular application of recursions is relative to the analysis of epidemiological stratified case-control studies. Smith et al. (1981) propose an algorithm for the ML estimates of the coefficients of a logistic regression and, since this procedure becomes computationally challenging for large strata, Krailo and Pike (1984) derive a recursive computation of the log-likelihood function of the conditional logit model and its derivatives. Further, Levin (1987) extend this result to multinomial outcomes. Other recursive solutions come from Item Response Theory (Hambleton and Swaminathan, 1986; Bartolucci et al., 2015), where Gustafsson (1980) shows a recursive structure for the conditional estimating equations of the Rasch model (Rasch, 1960).

In this regard, this work proposes a novel way to recursively compute the conditional likelihood function of the QE model. The main contribution is to extend the result of Krailo and Pike (1984) to dynamic models such as the QE model and its extensions recently proposed by Bartolucci and Nigro (2010), Bartolucci and Nigro (2012), and Bartolucci et al. (2018), whose computation would otherwise be not feasible with large time dimensions. Furthermore, I implement the proposed recursive algorithm in the `cquad` package[2], which provides software routines for the estimation, based on conditional maximum likelihood, of QE model and the related aforementioned contributions(Bartolucci and Pigini, 2017).

A Monte Carlo simulation shows how the recursive algorithm removes the computational burden of the QE model for large-T dataset. This fact should broaden the applicability of conditional inference for dynamic models in many economic applications. As an example, this work includes an empirical analysis concerning brand loyalty, where a QE model parameters are estimated exploiting the recursive algorithm.

The present work is organised as follows: Section 3.2 illustrates the main

[2]Available for the software R (`https://cran.r-project.org/package=cquad`) and Stata (`https://github.com/fravale/cquadr`).

theoretical aspects of the models and their conditional estimation, recalling some details presented in Chapter 1. Section 3.3 shows the proposed recursion for the QE model and its extensions. The performance of the algorithm in terms of computational time is evaluated by the Monte Carlo simulation in Section 3.4 and the analysis of brand loyalty to real data concerning yogurt purchases is reported in Section 3.5. Finally, Section 3.6 concludes.

3.2 Preliminaries

This section introduces at first the main theoretical aspects of the fixed-effects logit model and the QE model. Then, the Conditional Maximum Likelihood (CML) estimation for both models is presented.

3.2.1 Model assumptions

A general way to describe binary panel data model is the latent variable formulation, so that for an individual i, with $i = 1, \dots, n$, observed at the t-th time occasion, with $t = 1, \dots, T$, we model the binary response variable y_{it} as follows,

$$\begin{aligned} y_{it} &= \mathbb{1}\{y_{it}^* > 0\}, \\ y_{it}^* &= \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \end{aligned}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Moreover, the latent variable y_{it}^* is a linear index that includes a representation of the time-invariant individual unobserved heterogeneity α_i, a set of strictly exogenous time-varying covariates $\boldsymbol{x}_{it}$ related to the parameters $\boldsymbol{\beta}$, and an idiosyncratic error term ε_{it} that is assumed to follow a logistic distribution. Hereafter we will assumed α_i to be an individual specific intercept. This set of of assumptions allows us to derive the probability function for an observation, conditionally on the set of covariates $\boldsymbol{X}_i$ and on the individual effect, given by

$$p(y_{it}|\alpha_i, \boldsymbol{X}_{it}) = \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]}{1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})}.$$

The DL model is a straightforward extension of the static logit model, since it includes as a regressor the lagged dependent variable, $y_{i,t-1}$ along with exogenous covariates and the unobserved heterogeneity. Similar to the static case, we define the probability for the DL model as

$$p(y_{it}|\alpha_i, \boldsymbol{X}_{it}, y_{i0}, \dots, y_{i,t-1}) = \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma)]}{1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma)},$$

where the parameter γ measures the true state dependence and the conditioning set includes the individual outcome configuration, with the initial observation y_{i0} assumed to be known.

The QE model is of great interest, even because it closely resembles the DL model. In fact, the QE model allows us to include individual specific effects, a set of explanatory variables, and to measure true state dependence. The QE model directly formulates the probability for the individual outcome vector configuration $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})'$, so that

$$p(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp[y_{i+}\delta_i + (\sum_{t=1}^{T} y_{it}\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + y_{iT}(\phi + \boldsymbol{x}'_{iT}\boldsymbol{\eta}_2) + y_{i*}\psi]}{\sum_{\boldsymbol{z}} \exp[z_+\delta_i + (\sum_{t=1}^{T} z_t\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + z_T(\phi + \boldsymbol{x}'_{iT}\boldsymbol{\eta}_2) + z_{i*}\psi]},$$

where $y_{i*} = \sum_{t=1}^{T} y_{i,t-1}y_{it}$ and $y_{i+} = \sum_{t=1}^{T} y_{it}$ is defined as total score, δ_i denotes the unobserved heterogeneity, and ψ denotes the state dependence parameter. The vector $\boldsymbol{\eta}_1$ includes parameters of interest while ϕ and $\boldsymbol{\eta}_2$ are considered as nuisance parameters. The denominator is given by a sum extended to all the possible binary response vectors $\boldsymbol{z} = (z_1, ..., z_T)'$, where $z_+ = \sum_{t=1}^{T} z_t$ and $z_{i*} = y_{i0}z_1 + \sum_{t>1} z_{t-1}z_t$.

3.2.2 Conditional Inference

A seminal contribution in the field of conditional inference is the one of Chamberlain (1980), who derived a CML estimator for the parameters of the static fixed-effects logit model along the following lines. Consider the probability for the individual configuration $\boldsymbol{y}_i$, which is defined as

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{x}_i) = \prod_{t=1}^{T} p(y_{it}|\alpha_i, \boldsymbol{X}_{it}) = \frac{\exp[\alpha_i y_{i+} + (\sum_{t=1}^{T} y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\prod_{t=1}^{T}[1+\exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]}, \tag{3.1}$$

where the total score y_{i+} is proven to be a sufficient statistic for the nuisance parameter α_i. The point is to derive the probability of $\boldsymbol{y}_i$ conditional on the total score y_{i+}. Define now the probability of observing a given total score for the logit model, so that conditional on α_i and $\boldsymbol{X}_i$, we have

$$p(y_{i+}|\alpha_i, \boldsymbol{X}_i) = \frac{\sum_{\boldsymbol{z}:z_+=y_+} \exp(\alpha_i z_+)\exp[(\sum_{t=1}^{T} z_t \boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\prod_{t=1}^{T}[1+\exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]}, \tag{3.2}$$

where the numerator is given by the sum of all possible binary response vector $\boldsymbol{z} = (z_1, ..., z_T)'$ such that $z_+ = \sum_{t=1}^{T} z_t$ equals the total score, y_{i+}. The conditional probability of $\boldsymbol{y}_i$ given the total score, is defined as the ratio of the quantities defined in Equations (3.1) and (3.2), that is

$$\begin{aligned} & p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i+}) = \\ &= \frac{\exp(\alpha_i y_{i+})\exp[(\sum_{t=1}^{T} y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\prod_{t=1}^{T}[1+\exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]} \frac{\prod_{t=1}^{T}[1+\exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta})]}{\sum_{\boldsymbol{z}:z_+=y_+} \exp(\alpha_i z_+)\exp\left[\left(\sum_{t=1}^{T} z_t\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]} = \\ &= \frac{\exp[(\sum_{t=1}^{T} y_{it}\boldsymbol{x}_{it})'\boldsymbol{\beta}]}{\sum_{\boldsymbol{z}:z_+=y_+} \exp\left[\left(\sum_{t=1}^{T} z_t\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]} = p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}), \end{aligned} \tag{3.3}$$

where we end up with a conditional probability $p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+})$ that no longer depends on the nuisance parameter and will be the individual contribution to the conditional log-likelihood function

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbb{1}\{0 < y_{i+} < T\} \log p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}), \tag{3.4}$$

which can be maximised by a Newton-Raphson algorithm.

Unfortunately, it is not possible to extend this simple result to the DL model. Defining the probability of $\boldsymbol{y}_i$, we end up with

$$p(\boldsymbol{y}_i|\alpha_i, y_{i0}, \boldsymbol{X}_i) = \frac{\exp[y_{i+}\alpha_i + \left(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it}\right)' \boldsymbol{\beta} + y_{i*}\gamma]}{\prod_{t=1}^T [1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \gamma y_{i,t-1})]}$$

where the total score y_{i+} is not a sufficient statistic for α_i.

CML estimation of the DL model without covariates and $T = 3$ has been proposed by Chamberlain (1993). Later, Honoré and Kyriazidou (2000) proposed conditional estimation of a DL with covariates even though the identification is achieved relying on a kernel density function which rules out the use of discrete explanatory variables and reduces the rate of convergence of this estimator.

The QE model overcomes the drawbacks of the DL model. In particular, it admits the total score as a sufficient statistic for the nuisance parameters so that a conditional likelihood function can be built on the following probability

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp[(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + y_{iT}(\phi + \boldsymbol{x}_{iT}'\boldsymbol{\eta}_2) + y_{i*}\psi]}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp[(\sum_{t=1}^T z_t\boldsymbol{x}_{it})'\boldsymbol{\eta}_1 + z_T(\phi + \boldsymbol{x}_{iT}'\boldsymbol{\eta}_2) + z_{i*}\psi]}.$$

Parameters estimates can be easily obtained by a Newton-Raphson algorithm that maximises the likelihood function.

3.3 Proposed Methodology

A common shortcoming of the aforementioned models is the computational burden of the conditional likelihood function for a large time dimension of the panel dataset. This section introduces the result of Krailo and Pike (1984) for the static logit model and then shows the proposed methodology for the QE and its extensions.

3.3.1 Static Model

Consider the individual conditional probability in Equation (3.3), the denominator of that quantity is given by

$$\sum_{\boldsymbol{z}:z_+=y_+} \exp\left[\left(\sum_{t=1}^{T} z_t \boldsymbol{x}_{it}\right)' \boldsymbol{\beta}\right], \tag{3.5}$$

where the sum is extended to all the possible vector configurations $\boldsymbol{z}$ such that the sum of each vector, z_+, equals the total score y_{i+}. From the computational point of view, the sum above represents a limit for the applicability of the CML estimation. In fact, the number of such vectors $\boldsymbol{z}$ depends on the time dimension and on the the total score, so that when T is large their number dramatically increases. For example, for a response configuration $\boldsymbol{y}_i$ with $T = 4$ observations and total score $y_{i+} = 2$ we have $k = 6$ different vectors $\boldsymbol{z}$. We can collect them in a matrix $\boldsymbol{Z}$, with T rows and k columns, so that

$$\boldsymbol{y}_i = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \qquad \boldsymbol{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Table 3.1 shows the total number of column vectors in $\boldsymbol{Z}$ for a small grid of values in T and y_{i+}. When the number k of such vectors becomes too large, the computation of the term in Equation (3.5) is no longer feasible. However, as noted by Howard (1972), the computation of the sum in Equation (3.5) can be performed recursively. Define now a function $f_{t,s}(\cdot)$ indexed by the time dimension $t = 1, \dots, T$ and by the total score denoted $s = 0, \dots, y_{i+}$. This function is defined as

$$f_{t,s}(\boldsymbol{\phi}) = \sum_{\boldsymbol{z}:z_+=s} \exp\Big(\sum_{u=1}^{t} z_u \phi_u\Big),$$

where $\boldsymbol{z} = (z_1, \dots, z_t)'$ such that $\sum_{u=1}^{t} z_u = s$ and $\phi_u = \boldsymbol{x}'_{iu}\boldsymbol{\beta}$. Therefore, for $t = T$ and $s = y_{i+}$ this function equals Equation (3.5). Exploiting the

Table 3.1: Number of vectors in $\boldsymbol{Z}$

		T		
		12	18	25
y_{i+}	6	924	18564	177100
	9	220	48620	2042975
	12	1	18564	5200300

symmetric function properties of $f(\boldsymbol{\phi})$, it can be recursively computed in this way:

1. for $t = 1$, the process must be initialised; therefore, the function assumes values $f_{1,0}(\boldsymbol{\phi}) = 1$, $f_{1,1}(\boldsymbol{\phi}) = \exp(\phi_1)$, and 0 otherwise;

2. for $t = 2, \ldots, T$ it is possible compute the the function recursively as

$$f_{t,s}(\boldsymbol{\phi}) = f_{t-1,s}(\boldsymbol{\phi}) + f_{t-1,s-1}(\boldsymbol{\phi}) \exp(\phi_t), \quad s = 0, \ldots, t. \tag{3.6}$$

The recursive structure can be exploited for the computation the first- and the second-derivatives of this function. Those derivatives are part of the Score and the Hessian functions, respectively, that are involved in the Newton-Rhapson algorithm for the maximisation of the conditional log-likelihood function in Equation (3.4). Derivatives of the recursive function are reported in Appendix D.

3.3.2 Quadratic Exponential Model

As pointed out in the previous section, the aforementioned methodologies are suitable for the static logit model and their extension to dynamic models is not straightforward. This section shows how the proposed methodology works.

Consider the simplified version of the QE model described in Bartolucci and Pigini (2017). The conditional probability of $\boldsymbol{y}_i$, given $\boldsymbol{X}_i$, the total score y_{i+}, and the initial observation y_{i0}, is defined as

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_{t=1}^T y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta} + \sum_{t=1}^T y_{i,t-1}y_{it}\psi)}{\sum_{\boldsymbol{z}:z_+=\boldsymbol{y}_{i+}} \exp[\sum_{t=1}^T z_t\boldsymbol{x}'_{it}\boldsymbol{\beta} + (y_{i0}z_1 + \sum_{t=2}^T z_{i,t-1}z_{it})\psi]}. \tag{3.7}$$

In this case, given the time dimension, the total score, and the initial observation, the denominator in Equation (3.7) may be expressed as a function $g(\cdot)$, given by

$$g_{T,y_{i0},y_{i+}}(\boldsymbol{\phi},\psi) = g_{T,y_{i0},y_{i+},0}(\boldsymbol{\phi},\psi) + g_{T,y_{i0},y_{i+},1}(\boldsymbol{\phi},\psi). \tag{3.8}$$

In general, we define a single element of the sum in lhs of Equation (3.8) as

$$g_{t,a,s,v}(\boldsymbol{\phi},\psi) = \sum_{\boldsymbol{z}:z_+=s,z_t=v} \exp[\sum_{u=1}^t z_u\phi_u + (az_1 + \sum_{u=2}^t z_{u-1}z_u)\psi],$$

where the index v represent the t-th element of the vector $\boldsymbol{z}$, the initial observation is $a = 0, 1$, and following the notation above, $s = 1, \ldots, t$ and $t = 1, \ldots, T$ are the total score and the length of the vector configuration $\boldsymbol{y}_i$, respectively. The main difference with respect to the static case is that, further to the total score and the time dimension, we take into account not only the initial observation but also the last element v of the vector $\boldsymbol{z}$. Following the formulation above, this is how to compute $g_{t,a,s,v}(\boldsymbol{\phi},\psi)$ by a recursion:

1. for $t = 1$, the initialisation of the algorithm is done computing

$$g_{1,a,s,v}(\boldsymbol{\phi},\psi) = \begin{cases} 1, & s = v = 0, \\ \exp(\phi_1 + a\psi), & s = v = 1, \\ 0, & \text{otherwise}, \end{cases}$$

for $a = 0, 1$;

2. for $t = 2, \dots, T$, recursively compute the following quantities

$$g_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} 1, & s = 0,\ v = 0, \\ g_{t-1,a,s,0}(\boldsymbol{\phi}, \psi) + g_{t-1,a,s,1}(\boldsymbol{\phi}, \psi), & s = 1, \dots, t-1,\ v = 0, \\ g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi) \exp(\phi_t) + g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi) \exp(\phi_t + \psi), & s = 1, \dots, t,\ v = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.9}$$

First- and second-order derivatives required to perform the Newton-Rhapson algorithm are reported in Appendix D

3.3.3 Extensions

The proposed methodology can be easily built for the extensions of the QE model. In particular, we focus on the Pseudo Conditional Maximum Likelihood (PCML) estimator of the DL model proposed in Bartolucci and Nigro (2012) and the Modified QE model (Bartolucci et al., 2018), suitable to perform a test for state dependence.

Pseudo Conditional Maximum Likelihood Estimator

The PCML estimator is derived by approximating the DL model by a QE model. In this scenario, the conditional probability for $\boldsymbol{y}_i$ is given by

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_{t=1}^T y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta} - \sum_{t=1}^T \bar{q}_{it} y_{i,t-1}\gamma + \sum_{t=1}^T y_{it} y_{i,t-1}\gamma)}{\sum_{\boldsymbol{z}:z_+=\boldsymbol{y}_{i+}} \exp[\sum_{t=1}^T z_t \boldsymbol{x}'_{it}\boldsymbol{\beta} - (y_{i0}\bar{q}_{i1} + \sum_{t=2}^T \bar{q}_{it} z_{i,t-1})\gamma + (y_{i0}z_1 + \sum_{t=2}^T z_{it} z_{i,t-1})\gamma]}.$$

This model is built as an approximation of the DL model around $\alpha = \bar{\alpha}$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, and $\gamma = 0$ and the term $\bar{q}_{it}$ represents the probability of $y_{it} = 1$ for a logit model with parameters fixed as above. This terms are generated by an auxiliary regression and are related to the state dependence parameter. Moreover, define the terms $\phi_t = \boldsymbol{x}'_{it}\boldsymbol{\beta}$ and $\nu_t = -\bar{q}_{it}\gamma$.

Exploiting the same structure of Equation (3.8), the recursive computation of $g_{t,a,s,v}(\boldsymbol{\phi}, \gamma)$ can be performed as follows:

1. for $t = 1$ compute

$$g_{1,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} \exp(\nu_1 a), & s = v = 0 \\ \exp(\phi_1 + a(\nu_1 + \gamma)), & s = v = 1, \\ 0, & \text{otherwise,} \end{cases}$$

   for $a = 0, 1$;

2. for $t = 2, \ldots, T$, the recursion is

$$g_{t,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} 1, & s = 0,\ v = 0, \\ g_{t-1,a,s,0}(\boldsymbol{\phi}, \gamma) + g_{t-1,a,s,1}(\boldsymbol{\phi}, \gamma)\exp(\nu_t), & s = 1, \ldots, t-1,\ v = 0, \\ g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \gamma)\exp(\phi_t) + g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \gamma)\exp(\phi_t + \gamma + \nu_t), & s = 1, \ldots, t,\ v = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.10}$$

where, again, the first and second derivatives are reported in the Appendix D.

Modified Quadratic Exponential Model

The modified QE model closely resembles the simplified version of the QE, the main difference is that in this case the association rule between the outcome variable and its lag takes into account the pairs of consecutive observations that are equal, regardless if their value is 0 or 1.

Under the modified QE model, the conditional probability of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$ and y_{i0} is defined as

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} + \tilde{y}_{i*}\gamma)}{\sum_{\boldsymbol{z}:z_+=\boldsymbol{y}_{i+}} \exp\left(\sum_{t=1}^T z_t\boldsymbol{x}_{it}'\boldsymbol{\beta} + \tilde{z}_{i*}\gamma\right)}.$$

where $\tilde{y}_{i*} = \sum_t \mathbb{1}\{y_{it} = y_{i,t-1}\}$ and $\tilde{z}_{i*} = \mathbb{1}\{y_{i0} = z_{i1}\} + \sum_{t>1} \mathbb{1}\{z_{it} = z_{i,t-1}\}$.

For what concerns the computation of function $g_{t,a,s,v}(\boldsymbol{\phi}, \psi)$, the recursion is given by the formulation below:

1. for $t = 1$,

$$g_{1,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} \exp(\psi \mathbb{1}\{a = 0\}), & s = v = 0, \\ \exp(\phi_1 + \psi \mathbb{1}\{y_{i0} = 1\}), & s = v = 1, \\ 0, & \text{otherwise}, \end{cases}$$

   with $a = 0, 1$;

2. then, for $t = 2, \ldots, T$,

$$g_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} g_{t-1,a,s,0}(\boldsymbol{\phi}, \psi)\exp(\psi) + g_{t-1,a,s,1}(\boldsymbol{\phi}, \psi), & s = 0, \ldots, t-1,\ v = 0, \\ g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi), & s = 1, \ldots, t,\ v = 1, \\ 0, & \text{otherwise}. \end{cases} \tag{3.11}$$

Also in this case, the derivatives are reported in Appendix D.

3.4 Computational Complexity

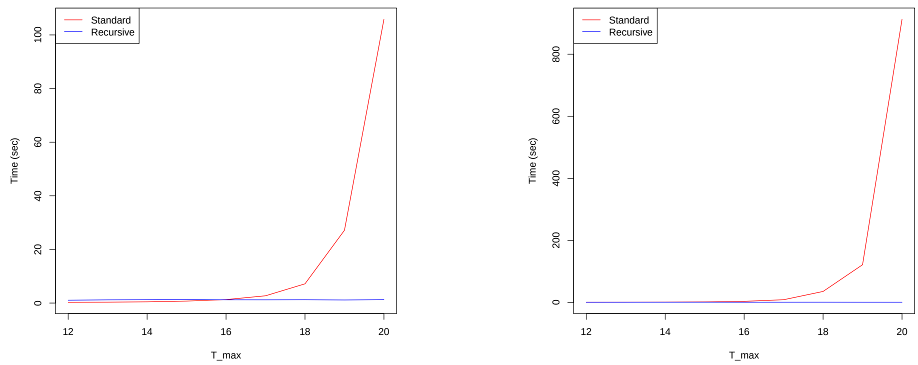
The relative advantage of the recursive algorithm in terms of computational time is evaluated in this section by a simple Monte Carlo simulation[3]. The simulation design is similar to the one of Honoré and Kyriazidou (2000). In this regard, data are generated from a DL model based on assuming

$$y_{i0} = \mathbb{1}\{\alpha_i + \beta x_{i0} + \varepsilon_{i0} > 0\},$$
$$y_{it} = \mathbb{1}\{\alpha_i + \gamma y_{i,t-1} + \beta x_{it} + \varepsilon_{it} > 0\},$$

for $i = 1 \ldots n$ and $t = 1 \ldots T$ beyond an initial observation, in $t = 0$. Further, x_{it} is an exogenous regressor generated from a Gaussian distribution with zero mean and variance $\pi^2/3$ and ε_{it} is a random variable following a standard logistic distribution. The parameter β is equal to 1 and the state dependence parameter γ is equal to 0.5. Individual effects are generated as $\alpha_i = \frac{1}{4}\sum_{t=0}^{3} x_{it}$. Finally, we consider $n = 250$ individuals observed at different T time occasions. In order simplify the interpretation of the results

[3]Experiments were run on a computer with n. 2 Intel Xeon CPU E5-2640 v4 2.40GHz, 250GB of RAM, running Debian GNU/Linux "bullseye"/sid as operating system.

Figure 3.1: CPU time comparison of algebraic and recursive computation



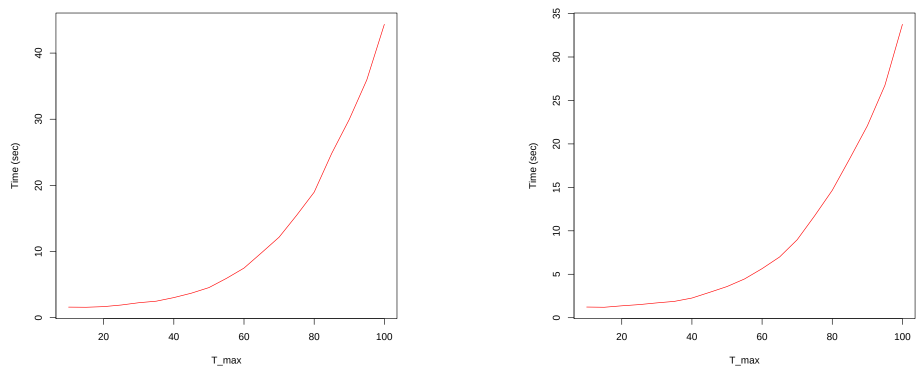
(a) QE (b) PCML

we are are going to consider a dataset given by 249 subjects observed $T = 5$ time occasions and the last one, the 250$-$th, observed T_{max} time occasions.

The basic design illustrated above is adopted for two different experiments. The first one aims to compare the time required for the computation of the likelihood function with both the standard algebraic operation and using the recursive algorithm. In this case, T_{max} ranges in $\{12, \dots, 20\}$. With the second set of simulations we want to explore how the CPU time required by the algorithm varies according to the time dimension of the panel. In this case we evaluate T_{max} in a grid between 10 and 100 with steps of 5. The number of Monte Carlo replications is 50 for the first experiment and 25 for the second.

Figure 3.1 represents the average computational time required for the likelihood maximisation of the simplified QE model (3.1a) and the PCML estimator (3.1b). For the two models, each subfigure reports the time required by the algebraic operations ("Standard") and the recursion ("Recursive"), in which we can see that the former exhibits an exponential pace as T_{max} approaches 18 while the latter remains stable. The time taken by the algebraic computation of the QE grows up to 100 second while the PCML routine takes more than 13 minutes, on average, for $T = 20$. For what concerns the algorithm, the computational time is always lower than one second for all

Figure 3.2: CPU time of the recursion in large time dimensions



(a) QE (b) PCML

the time dimensions considered.

The derivation of the computational complexity of the proposed algorithm is straightforward. Consider the structure of the recursion in Equation (3.9), basic operations such as sums and products are computed for each value of the total score s in $1, \dots, t$, and repeated for t varying in $1, \dots, T$ so that the potential leading order of operations is $O(T?)$ where, following the notation of Knuth, $T?$ denotes the termial function of T, namely the sum of all positive integers less than or equal to T (Knuth, 1997).

The second set of simulations allows us to evaluate the CPU time required by the recursion for larger time dimensions as reported in Figure 3.2. The average CPU time taken by the maximisation, reported in lines, is obviously increasing in T_{max} and follows a path that is coherent with the complexity derived above.

3.5 Application: Brand Loyalty

The proposed methodology is applied to real data concerning brand loyalty. The analysis here performed aims to replicate of the one in Hsiao (2014), Section 7.5.5.2, concerning consumers loyalty to two different dominant yogurt brands. Differently from the original work, the present application is based on a sample dataset provided by A.C. Nielsen[4] that contains information about yogurt purchases made by individuals observed for a period of about two years. This dataset has already been used by Jain et al. (1994). Data are about purchases of four brands: Yoplait, Dannon, Nordica and Weight. As in Hsiao, we keep observations of purchases of the two brands with the largest market share in the dataset, namely Dannon and Yoplait. The dependent variable is set to 1 when a consumer choose the brand "Dannon". We also include two exogenous explanatory variables. The first one is the prices log-difference ("Price") for the two products. The second one is a categorical variable ("Featured") and is built as the difference of two dummy variables relative to the brands, which record whether a brand is advertised in newspapers. We expect that a small relative price and advertisement should have a positive impact on the demand of the goods and then on the probability of a purchase for a brand. The actual sample consists of an unbalanced panel of 100 consumers, for a total of 1,788 observations. Time occasions differ among individuals and range between a minimum of $T_{min} = 1$ and a maximum of $T_{max} = 161$.

The models here considered are extremely useful to represent consumers behaviour and their loyalty to a brand (Chintagunta et al., 2001). First of all, from a statistical perspective, it is straightforward to represent brand choices as a binary variable being 1 or 0 according to whether a product is chosen or not. Secondly, the panel structure of the data allows us to take into account some individual time-constant unobserved characteristics. Since consumers are observed for a relatively short span of time in which their purchases are recorded, from a theoretical perspective it is reasonable to consider some individual unobservables being time-invariant. Finally, consumer habits stickiness require to be taken into account by a dynamic specification of the model that should identify true state dependence. Hsiao shows different model specifications but only the DL model properly includes individual fixed-effects and accounts for state dependence. With respect to the original work, where the DL is estimated following Honoré and Kyriazidou (2000),

[4]Data are publicly available in the `R` package `Ecdat` (`https://cran.r-project.org/web/packages/Ecdat/index.html`)

Table 3.2: Estimation results

	Pooled	CML	HK	**QE**	**PCML**
ϕ	4.634	1.715	1.873	2.118	2.326
	(0.183)	(0.317)	(0.193)	(0.221)	(0.389)
β_p	-3.373	-3.565	-2.902	-3.264	-3.390
	(0.384)	(0.771)	(0.564)	(0.514)	(0.702)
β_f	0.362	0.739	0.120	0.440	0.723
	(0.275)	(0.490)	(0.341)	(0.317)	(0.438)
Time (sec.)	< 1	1.304	2.236	28.264	27.576

[1] Standard errors in parentheses

the proposed algorithm allows us to extend the analysis to a wider set of estimators such as the CML estimator for the QE model and the PCML for the DL model. We consider different estimation techniques of the DL model:

$$Dannon_{it} = \mathbb{1}\{\alpha_i + \phi Dannon_{i,t-1} + \beta_p Price_{it} + \beta_f Featured_{it} + \varepsilon_{it} > 0\}$$

where ϕ is the state dependence parameter and where β_p and β_f are the regression parameters. Table 3.2 reports the estimated coefficients for the pooled model (Pooled) and the CML estimator of a static logit model with fixed effects where the lag is treated as exogenous . Further we can find the Honoré and Kyriazidou (2000) estimator (HK), the estimation of the simplified QE model and the PCML estimator.

For what concerns the models in Table 3.2, estimated coefficients differ across the proposed model specifications. In general, all the signs are coherent with economic theory so that an increase of the relative price decreases the probability of a purchase for the brand "Dannon". The variable "Featured" seems to be not statistically significant. Furthermore, it is interesting to see how the estimated level of state dependence is sensitive to the specification of the unobservables. In the pooled model the state dependence estimated parameter is the largest among all the considered specifications. Including heterogeneity dramatically decreases the estimated coefficient, which slightly differs between the HK and the PCML estimators due to the different estimation strategies. Anyway, the Pooled and the CML models are likely to provides bias estimates because the first ignores the potential heterogeneity and in both models the dynamic specification is not accounted for. Finally, we observe that the computational time of the QE and PCML models are about twenty times larger with respect to the static CML because of the larger complexity of the dynamic models considered with respect to the static one.

It is worth stressing that there is no way to compare the computational time required by the algorithm for the QE and the PCML models and their standard matrix computation since, in the latter case, the estimation in not feasible due to the large numbers of time occasions.

3.6 Conclusions

This work provides a novel way to compute the conditional likelihood functions of the QE model and its extensions. By means of a recursive algorithm, standard computational burden of algebra calculations is overcome for applications that involve longitudinal panel dataset with large time dimensions. This novelty should be relevant in presence of unbalanced panel dataset which makes conditional inference more appealing since the ML estimator and its corrections are hampered by this kind of dataset. A Monte Carlo experiment shows how the proposed algorithm outperforms standard computation and it should enlarge the applicability of the considered models. In this regard, an application to real data concerning brand loyalty has been proposed with the additional interest of showing a set of results that would have been otherwise impossible, providing practitioners with a wider range of estimation tools for empirical analysis.

Conclusion

This work discusses real data problems concerning the estimation of binary panel data models with fixed-effects, focusing on the conditional inference appraoch. I focus on two aspects that limit the applicability of these techniques: endogenous selection mechanisms and the computational burden of the conditional likelihood function of the QE models. This task is accomplished by three different studies summarized in three chapters.

A systematic literature review highlights the validity of the conditional inference approach. By means of an extensive Monte Carlo study, I show how these techniques outperform the alternative bias-corrected estimators for panel datasets charachterised by a small number of time time occasions.

The computational issue is dealt with by exploiting recursive algorithms that rule out the limits of the QE models for some economic applications. The proposed algortithm allows practiotioners to exploit conditional inference for problems that involve *long* panel dataset, widening the set of suitable applications that would have been otherwise infeasible. This is confirmed by both a Monte Carlo simulation and a real data application.

The results concerning the issue of endogenous selection mechanisms are far from being exhaustive. I propose an approximanting model, estimated by a PCML, that allows to detect wherther the selection is endogenous or not. Despite the fact that the test always attains the nominal size and exhibits good power properties, it is not clear how to deal with the endogeneity of the selection mechanisms in estimation. The proposed model is not able to provide a consistent estimator of the paramters and further research is needed in order to characterise the approximation bias.

Finally, the conditional inference framework requires additional investigations. In fact, the estimation of marginal effects, which is crucial in empirical analysis, is still an open issue. Although CML provides a consistent estimator for regression parameters, it is crucial to understand the proper estimator for individual intercepts which are required for the computation of marginal effects. Moreover, sufficient statistics for the incidental parameters cannot be derived for a large number of models, as well as for a generalisation to multivariate models, which would be key to properly deal with phenomena like selection mechanisms. An additional extension of the whole framework could be a generalisation to models that account for time-varying unobserved heterogeneity even if this field is still unexplored.

References

Ahn, H. and Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58:3–29.

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

Alessie, R., Hochguertel, S., and Van Soest, A. (2004). Ownership of stocks and mutual funds: A panel data analysis. *The review of Economics and Statistics*, 86:783–796.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. *Econometrica*, 46:1193–1205.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B*, 32:283–301.

Andersen, E. B. (1971). The asymptotic distribution of conditional likelihood ratio tests. *Journal of the American Statistical Association*, 66:630–633.

Angelini, V., Brugiavini, A., and Weber, G. (2009). Ageing and unused capacity in europe: is there an early retirement trap? *Economic Policy*, 24:463–508.

Arabmazar, A. and Schmidt, P. (1982). An investigation of the robustness of the tobit estimator to non-normality. *Econometrica*, 50:1055–1063.

Arellano, M. (2003). Discrete choices with panel data. *Investigaciones Económicas*, 27:423–458.

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

Arellano, M., Bover, O., and Labeaga, J. M. (1999). Autoregressive models with sample selectivity for panel data. In Hsiao C, Lahiri K, L. L. and MH, P., editors, *Analysis of Panels and Limited Dependent Variable Models*, pages 23–48. Cambridge University Press, Cambridge, UK.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, 43:381.

Arellano, M. and Hahn, J. (2016). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. *Global Economic Review*, 45:251–274.

Bartolucci, F., Bacci, S., and Gnaldi, M. (2015). *Statistical analysis of questionnaires: A unified approach based on R and Stata*. Chapman and Hall/CRC.

Bartolucci, F., Bellio, R., Salvan, A., and Sartori, N. (2016). Modified profile likelihood for fixed-effects panel data models. *Econometric Reviews*, 35:1271–1289.

Bartolucci, F. and Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. *Econometrica*, 78:719–733.

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

Bartolucci, F., Nigro, V., and Pigini, C. (2018). Testing for state dependence in binary panel data with individual covariates by a modified quadratic exponential model. *Econometric Reviews*, 37:61–88.

Bartolucci, F. and Pennoni, F. (2007). On the approximation of the quadratic exponential distribution in a latent variable context. *Biometrika*, 94:745–754.

Bartolucci, F. and Pigini, C. (2017). cquad: An r and stata package for conditional maximum likelihood estimation of dynamic binary panel data models. *Journal of Statistical Software*, 78.

Bartolucci, F. and Pigini, C. (2019). Partial effects estimation for fixed-effects logit panel data models.

Bester, C. A. and Hansen, C. (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics*, 27:131–148.

Bettin, G. and Lucchetti, R. (2016). Steady streams and sudden bursts: persistence patterns in remittance decisions. *Journal of Population Economics*, 29:263–292.

Blundell, R. and Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357.

Brown, S., Ghosh, P., and Taylor, K. (2014). The existence and persistence of household financial hardship: A bayesian multivariate dynamic logit framework. *Journal of Banking & Finance*, 46:285–298.

Butler, J. S. and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50:761–764.

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

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

Celidoni, M., Dal Bianco, C., and Weber, G. (2017). Retirement and cognitive decline. A longitudinal analysis using SHARE data. *Journal of health economics*, 56:113–125.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1).

Chamberlain, G. (1993). Feedback in panel data medels. Technical report, Harvard-Institute of Economic Research.

Chintagunta, P., Kyriazidou, E., and Perktold, J. (2001). Panel data analysis of household brand choices. *Journal of Econometrics*, 103:111–153.

Coe, N. B. and Zamarro, G. (2011). Retirement effects on health in europe. *Journal of health economics*, 30:77–86.

Comi, S., Cottini, E., and Lucifora, C. (2019). The effect of retirement on social relationship: new evidence from SHARE data.

Cox, D. R. (1972a). The analysis of multivariate binary data. *Applied statistics*, 21:113–120.

Cox, D. R. (1972b). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, 34:187–202.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B*, 49:1–18.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82:991–1030.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of health economics*, 43:1–12.

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

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192:291–312.

Fernández-Val, I. and Weidner, M. (2018). Fixed effects estimation of large-t panel data models. *Annual Review of Economics*, 10:109–138.

Gayle, G.-L. and Viauroux, C. (2007). Root-n consistent semiparametric estimators of a dynamic panel-sample-selection model. *Journal of Econometrics*, 141:179–212.

Giarda, E. (2013). Persistency of financial distress amongst italian households: Evidence from dynamic models for binary panel data. *Journal of Banking & Finance*, 37:3425–3434.

Greene, W. H. (2011). Econometric analysis, international edition, 7th edition.

Gustafsson, J.-E. (1980). A solution of the conditional estimation problem for long tests in the rasch model for dichotomous items. *Educational and Psychological Measurement*, 40:377–385.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27:1152–1191.

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

Hambleton, R. K. and Swaminathan, H. (1986). *Item response theory: Principles and applications*. Springer Science & Business Media.

Hausman, J. A. (1975). An instrumental variable approach to full information estimators for linear and certain nonlinear econometric models. *Econometrica*, 43:727–738.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, 42:679–694.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 475–492. NBER.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46:931–959.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47:153–161.

Heckman, J. J. (1981a). Heterogeneity and state dependence. In *Studies in labor markets*, pages 91–140. University of Chicago Press.

Heckman, J. J. (1981b). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some monte carlo evidence. *Structural Analysis of Discrete Data with Economic Applications, Eds. Manski CF and McFadden D., MIT Press, Cambridge MA*.

Heckman, J. J. and Borjas, G. J. (1980). Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica*, 47:247–283.

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

Howard, S. (1972). Remark on the paper by Cox, D. R. (1972): Regression methods and life tables. *Journal of the Royal Statistical Society: Series B*, 34:187–202.

Hsiao, C. (2014). *Analysis of panel data*. Cambridge university press.

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

Jain, D. C., Vilcassim, N. J., and Chintagunta, P. K. (1994). A random-coefficients logit brand-choice model applied to panel data. *Journal of Business & Economic Statistics*, 12:317–328.

Knuth, D. E. (1997). *The art of computer programming*, volume 3. Pearson Education.

Krailo, M. and Pike, M. (1984). Algorithm AS 196: conditional multivariate logistic analysis of stratified case-control studies. *Journal of the Royal Statistical Society. Series C*, 33:95–103.

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica*, 65:1335–1364.

Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. *The Review of Economic Studies*, 68:543–572.

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

Lee, L. F. (1981). Simultaneous equations models with discrete and censored variables. In *Econometric analysis of discrete data*.

Levin, B. (1987). Conditional likelihood analysis in stratum-matched retrospective studies with polytomous disease states. *Communications in Statistics-Simulation and Computation*, 16:699–718.

Lin, W. and Wooldridge, J. M. (2017). Binary and fractional response models with continuous and binary endogenous explanatory variables.

Lin, W. and Wooldridge, J. M. (2019). Testing and correcting for endogeneity in nonlinear unobserved effects models. In *Panel Data Econometrics*, pages 21–43. Elsevier.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. John Wiley & Sons.

Mazzonna, F. and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56:691–710.

Mazzonna, F. and Peracchi, F. (2017). Unhealthy retirement? *Journal of Human Resources*, 52:128–151.

McCullagh, P. and Nelder, J. A. (1989). Generalized linear models, no. 37 in monograph on statistics and applied probability.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142.

Monfardini, C. and Radice, R. (2008). Testing exogeneity in the bivariate probit model: A monte carlo study. *oxford Bulletin of Economics and Statistics*, 70:271–282.

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

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, 36:231–250.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32.

Nicoletti, C. and Peracchi, F. (2001). Two-step estimation of binary response models with sample selection. In *first British Household Panel Survey research conference held in Colchester*. Citeseer.

Papke, L. E. and Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145:121–133.

Pigini, C. (2015). Bivariate non-normality in the sample selection model. *Journal of Econometric Methods*, 4:123–144.

Pigini, C., Presbitero, A. F., and Zazzaro, A. (2016). State dependence in access to credit. *Journal of Financial Stability*, 27:17–34.

Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of economic surveys*, 14:53–68.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. *Copenhagen, Denmark: Danish Institute for Educational Research.*

Rivers, D. and Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39:347–366.

Rochina-Barrachina, M. E. (1999). A new estimator for panel data sample selection models. *Annales d'Economie et de Statistique*, pages 153–181.

Semykina, A. and Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157:375–380.

Semykina, A. and Wooldridge, J. M. (2013). Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics*, 28:47–61.

Semykina, A. and Wooldridge, J. M. (2018). Binary response panel data models with sample selection and self-selection. *Journal of Applied Econometrics*, 33:179–197.

Severini, T. A. (1998). An approximation to the modified profile likelihood function. *Biometrika*, 85:403–411.

Smith, P., Pike, M., Hill, A., Breslow, N., and Day, N. (1981). Algorithm AS 162: Multivariate conditional logistic analysis of stratum-matched case-control studies. *Journal of the Royal Statistical Society. Series C*, 30:190–197.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, 27:531–543.

Van de Ven, W. P. and Van Praag, B. M. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics*, 17:229–252.

Vella, F. (1992). Simple tests for sample selection bias in censored and discrete choice models. *Journal of Applied Econometrics*, 7:413–421.

Vella, F. and Verbeek, M. (1999). Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics*, 90:239–263.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68:115–132.

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

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, 182:226–234.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50:420–445.

Appendices

Appendix A

Additional simulation results

This appendix includes the results of the full set of simulation designs adopted in Section 1.3. Each table reports the mean bias, the median bias, the median absolute error, and the root mean square error for the the estimated coefficients of β and γ. Table A.1 recalls the notation used in Section 1.3. Tables are ordered such that the data generating process is given by a value of γ increasing from 0 to 2.

Table A.1: Parameters and estimators

Parameters		
β	Regression parameter	
γ	State dependence parameter	
Estimators		Reference
INF	Infeasible Likelihood	
ML	Maximum Likelihood	
HK	DL conditional estimator	Honoré and Kyriazidou (2000)
PCML	Pseudo Conditional ML	Bartolucci and Nigro (2012)
MPL	Modified Profile Likelihood	Bartolucci et al. (2016)
MML	Modifed ML	Carro (2007)
SPJ	Split-Panel Jackknife	Dhaene and Jochmans (2015)

Table A.2: Simulation results under the design based on Equations (1.23) and (1.24) with $\gamma = 0$, $\beta = 1$

Results for $\hat{\beta}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.009	0.806	0.100	0.023	-0.025	-0.056	-	0.088	0.848	0.275	0.133	0.150	0.094	-
	4	0.017	0.578	0.031	0.016	0.003	0.005	-	0.076	0.600	0.136	0.091	0.071	0.068	-
	6	0.007	0.330	0.010	0.010	0.023	0.024	-0.167	0.057	0.346	0.088	0.069	0.069	0.070	0.222
	8	-0.000	0.215	0.005	0.002	0.012	0.012	-0.126	0.048	0.227	0.065	0.054	0.055	0.055	0.151
	12	0.001	0.128	0.002	0.000	0.005	0.005	-0.062	0.038	0.137	0.048	0.042	0.042	0.042	0.078
500	3	0.004	0.771	0.041	0.004	-0.083	-0.066	-	0.065	0.789	0.173	0.084	0.108	0.081	-
	4	0.001	0.548	0.007	-0.000	-0.013	-0.006	-	0.051	0.562	0.098	0.069	0.083	0.054	-
	6	0.005	0.324	0.008	0.006	0.020	0.020	-0.142	0.040	0.331	0.061	0.047	0.049	0.049	0.174
	8	0.002	0.216	0.004	0.002	0.013	0.013	-0.124	0.034	0.222	0.050	0.039	0.041	0.041	0.137
	12	0.006	0.135	0.006	0.006	0.011	0.011	-0.055	0.030	0.141	0.038	0.034	0.035	0.035	0.065
1000	3	0.011	0.799	0.033	0.020	-0.021	-0.053	-	0.043	0.810	0.117	0.065	0.100	0.063	-
	4	0.001	0.556	0.010	0.005	-0.004	-0.000	-	0.035	0.562	0.072	0.047	0.037	0.036	-
	6	0.000	0.315	0.003	0.000	0.014	0.015	-0.143	0.029	0.319	0.046	0.034	0.036	0.036	0.157
	8	0.000	0.214	0.003	0.000	0.011	0.011	-0.121	0.024	0.217	0.035	0.028	0.030	0.030	0.127
	12	0.001	0.128	0.001	0.001	0.005	0.005	-0.059	0.020	0.130	0.025	0.021	0.021	0.021	0.064

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.009	0.780	0.077	0.014	-0.052	-0.058	-	0.056	0.780	0.160	0.090	0.076	0.065	-
	4	0.011	0.570	0.021	0.007	0.001	0.004	-	0.051	0.570	0.077	0.051	0.042	0.042	-
	6	0.006	0.327	0.009	0.008	0.024	0.024	-0.161	0.038	0.327	0.054	0.047	0.046	0.046	0.164
	8	-0.002	0.213	0.000	0.001	0.010	0.010	-0.126	0.032	0.213	0.048	0.038	0.038	0.038	0.126
	12	-0.002	0.126	0.000	-0.001	0.004	0.004	-0.064	0.025	0.126	0.033	0.029	0.029	0.029	0.064
500	3	0.006	0.791	0.011	0.010	-0.072	-0.063	-	0.044	0.791	0.111	0.066	0.072	0.063	-
	4	0.001	0.544	0.001	-0.001	-0.007	-0.005	-	0.034	0.544	0.065	0.044	0.035	0.035	-
	6	0.006	0.321	0.005	0.004	0.019	0.019	-0.142	0.026	0.321	0.040	0.031	0.032	0.033	0.142
	8	0.002	0.215	0.002	0.001	0.012	0.012	-0.126	0.024	0.215	0.035	0.026	0.027	0.027	0.126
	12	0.008	0.134	0.005	0.006	0.011	0.011	-0.056	0.020	0.134	0.024	0.021	0.022	0.021	0.056
1000	3	0.014	0.804	0.027	0.017	-0.042	-0.054	-	0.026	0.804	0.063	0.042	0.060	0.054	-
	4	0.001	0.554	0.012	0.004	-0.004	0.000	-	0.024	0.554	0.046	0.031	0.025	0.025	-
	6	0.001	0.315	0.001	0.000	0.014	0.015	-0.142	0.019	0.315	0.030	0.021	0.023	0.024	0.142
	8	0.000	0.214	0.002	0.000	0.012	0.012	-0.121	0.016	0.214	0.024	0.019	0.020	0.020	0.121
	12	0.000	0.128	0.001	0.001	0.006	0.006	-0.059	0.014	0.128	0.018	0.014	0.014	0.014	0.059

Table A.3: Simulation results under the design based on Equations (1.23) and (1.24) with $\gamma = 0$, $\beta = 1$

Results for $\hat{\gamma}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.008	-2.729	0.000	0.039	0.055	-0.350	-	0.160	2.796	0.539	0.407	0.350	0.420	-
	4	0.009	-1.695	0.018	0.035	-0.088	-0.201	-	0.115	1.733	0.321	0.263	0.220	0.285	-
	6	-0.005	-0.974	0.007	-0.001	-0.087	-0.109	0.837	0.104	0.996	0.202	0.172	0.185	0.194	0.908
	8	-0.005	-0.679	-0.001	-0.006	-0.056	-0.063	0.345	0.090	0.700	0.158	0.150	0.158	0.159	0.404
	12	-0.001	-0.415	-0.001	-0.000	-0.022	-0.024	0.122	0.068	0.431	0.113	0.110	0.111	0.111	0.179
500	3	-0.005	-2.759	-0.016	-0.012	0.003	-0.384	-	0.101	2.794	0.378	0.287	0.173	0.418	-
	4	0.001	-1.695	0.005	0.014	-0.109	-0.222	-	0.092	1.716	0.240	0.199	0.188	0.270	-
	6	-0.002	-0.975	-0.000	-0.002	-0.088	-0.110	0.794	0.074	0.987	0.152	0.132	0.153	0.166	0.831
	8	0.001	-0.672	0.001	0.001	-0.050	-0.057	0.341	0.063	0.681	0.112	0.102	0.111	0.114	0.369
	12	0.009	-0.401	0.000	0.012	-0.008	-0.011	0.132	0.054	0.411	0.081	0.084	0.083	0.083	0.159
1000	3	-0.006	-2.748	-0.018	-0.007	-0.037	-0.379	-	0.080	2.764	0.260	0.183	0.190	0.393	-
	4	0.001	-1.722	-0.012	-0.009	-0.130	-0.240	-	0.064	1.732	0.168	0.141	0.177	0.265	-
	6	0.003	-0.970	0.001	0.001	-0.086	-0.108	0.783	0.048	0.976	0.103	0.089	0.121	0.136	0.802
	8	-0.002	-0.674	-0.001	-0.002	-0.052	-0.060	0.337	0.043	0.679	0.082	0.072	0.088	0.092	0.352
	12	-0.001	-0.415	0.001	-0.001	-0.022	-0.024	0.120	0.035	0.419	0.058	0.054	0.058	0.059	0.135

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	-0.008	-2.703	0.009	0.061	0.017	-0.345	-	0.123	2.703	0.367	0.263	0.184	0.345	-
	4	0.001	-1.680	0.042	0.051	-0.082	-0.197	-	0.079	1.680	0.207	0.173	0.149	0.200	-
	6	-0.010	-0.976	0.013	0.009	-0.084	-0.104	0.824	0.073	0.976	0.140	0.115	0.118	0.133	0.824
	8	-0.008	-0.669	0.018	-0.000	-0.045	-0.052	0.344	0.058	0.669	0.107	0.105	0.105	0.098	0.344
	12	-0.002	-0.416	-0.006	-0.000	-0.021	-0.024	0.114	0.046	0.416	0.075	0.072	0.071	0.071	0.124
500	3	-0.011	-2.759	-0.050	-0.003	0.000	-0.396	-	0.074	2.759	0.241	0.209	0.097	0.396	-
	4	0.001	-1.683	0.006	0.022	-0.107	-0.216	-	0.064	1.683	0.176	0.146	0.125	0.216	-
	6	-0.003	-0.978	-0.002	-0.004	-0.088	-0.112	0.788	0.050	0.978	0.104	0.089	0.113	0.122	0.788
	8	0.001	-0.668	0.003	0.003	-0.046	-0.054	0.336	0.040	0.668	0.078	0.068	0.076	0.077	0.336
	12	0.007	-0.402	-0.001	0.010	-0.009	-0.013	0.131	0.041	0.402	0.051	0.050	0.049	0.051	0.131
1000	3	-0.002	-2.706	0.004	0.001	-0.000	-0.376	-	0.059	2.706	0.193	0.139	0.075	0.376	-
	4	0.001	-1.716	-0.012	-0.005	-0.126	-0.240	-	0.046	1.716	0.118	0.098	0.131	0.240	-
	6	0.001	-0.969	0.001	-0.001	-0.086	-0.107	0.778	0.033	0.969	0.069	0.061	0.090	0.107	0.778
	8	-0.003	-0.669	-0.000	0.001	-0.049	-0.057	0.338	0.029	0.669	0.056	0.048	0.059	0.061	0.338
	12	-0.001	-0.416	-0.002	-0.001	-0.022	-0.024	0.120	0.023	0.416	0.040	0.033	0.040	0.039	0.120

Table A.4: Design: $\gamma = 0.25$, $\beta = 1$

Results for $\hat{\beta}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.010	0.802	0.116	0.017	-0.025	-0.057	-	0.089	0.844	0.318	0.127	0.159	0.088	-
	4	0.010	0.598	0.056	0.028	0.016	0.016	-	0.077	0.626	0.171	0.106	0.080	0.080	-
	6	0.004	0.314	0.007	-0.000	0.016	0.016	-0.168	0.052	0.327	0.082	0.062	0.062	0.062	0.226
	8	0.001	0.215	0.006	0.002	0.014	0.013	-0.124	0.047	0.226	0.064	0.051	0.053	0.053	0.149
	12	0.001	0.129	0.004	0.001	0.007	0.006	-0.061	0.040	0.139	0.052	0.044	0.045	0.045	0.079
500	3	0.009	0.789	0.096	0.012	-0.063	-0.056	-	0.054	0.808	0.199	0.087	0.100	0.076	-
	4	0.005	0.561	0.030	0.009	0.006	0.005	-	0.049	0.573	0.108	0.066	0.052	0.051	-
	6	0.004	0.320	0.014	0.004	0.020	0.019	-0.155	0.041	0.328	0.063	0.047	0.049	0.049	0.187
	8	0.002	0.216	0.008	0.003	0.014	0.014	-0.120	0.034	0.222	0.049	0.038	0.040	0.040	0.134
	12	0.003	0.132	0.008	0.004	0.009	0.009	-0.053	0.027	0.137	0.037	0.029	0.030	0.030	0.063
1000	3	0.001	0.775	0.033	0.006	-0.031	-0.057	-	0.044	0.787	0.122	0.068	0.092	0.069	-
	4	-0.001	0.542	0.010	-0.002	-0.003	-0.002	-	0.036	0.548	0.074	0.047	0.037	0.037	-
	6	0.000	0.313	0.006	-0.000	0.016	0.016	-0.143	0.028	0.317	0.045	0.035	0.037	0.037	0.161
	8	0.001	0.214	0.005	0.001	0.013	0.013	-0.116	0.024	0.217	0.035	0.026	0.029	0.029	0.123
	12	0.000	0.128	0.002	0.000	0.006	0.005	-0.057	0.019	0.130	0.025	0.020	0.021	0.021	0.062

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.008	0.773	0.058	0.004	-0.051	-0.057	-	0.056	0.773	0.145	0.080	0.073	0.062	-
	4	0.004	0.590	0.039	0.023	0.014	0.015	-	0.055	0.590	0.094	0.075	0.058	0.059	-
	6	-0.001	0.302	0.008	-0.007	0.007	0.007	-0.167	0.033	0.302	0.054	0.044	0.041	0.041	0.168
	8	0.002	0.211	0.001	-0.001	0.011	0.011	-0.126	0.031	0.211	0.040	0.033	0.034	0.034	0.127
	12	0.001	0.129	0.004	0.001	0.007	0.007	-0.060	0.026	0.129	0.034	0.027	0.028	0.028	0.061
500	3	0.003	0.773	0.078	0.005	-0.066	-0.062	-	0.037	0.773	0.118	0.056	0.071	0.063	-
	4	0.003	0.556	0.024	0.008	0.005	0.006	-	0.033	0.556	0.069	0.044	0.035	0.034	-
	6	0.002	0.316	0.008	-0.001	0.016	0.016	-0.158	0.028	0.316	0.044	0.029	0.031	0.031	0.159
	8	0.003	0.217	0.007	0.004	0.016	0.016	-0.120	0.024	0.217	0.032	0.026	0.028	0.028	0.120
	12	-0.000	0.128	0.005	0.001	0.007	0.006	-0.054	0.017	0.128	0.025	0.019	0.019	0.018	0.054
1000	3	-0.002	0.771	0.029	0.001	-0.045	-0.056	-	0.033	0.771	0.076	0.045	0.055	0.057	-
	4	-0.002	0.537	0.005	-0.003	-0.004	-0.003	-	0.024	0.537	0.047	0.031	0.025	0.025	-
	6	-0.001	0.311	0.005	-0.002	0.015	0.015	-0.145	0.020	0.311	0.035	0.021	0.025	0.024	0.145
	8	-0.000	0.213	0.003	0.000	0.012	0.012	-0.117	0.016	0.213	0.024	0.018	0.020	0.020	0.117
	12	-0.000	0.128	0.003	0.000	0.006	0.006	-0.058	0.013	0.128	0.017	0.014	0.014	0.014	0.058

Table A.5: Simulation results under the design based on Equations (1.23) and (1.24) with $\gamma = 0.25$, $\beta = 1$

Results for $\hat{\gamma}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.011	-2.675	0.008	0.061	-0.085	-0.447	-	0.151	2.760	0.599	0.468	0.317	0.510	-
	4	0.006	-1.686	-0.069	-0.018	-0.182	-0.290	-	0.132	1.722	0.350	0.272	0.275	0.354	-
	6	-0.007	-0.931	-0.020	0.008	-0.093	-0.119	0.790	0.098	0.954	0.212	0.176	0.190	0.203	0.869
	8	0.003	-0.648	-0.022	0.005	-0.051	-0.061	0.329	0.084	0.665	0.155	0.135	0.141	0.144	0.379
	12	-0.002	-0.404	-0.025	-0.001	-0.024	-0.027	0.107	0.068	0.420	0.115	0.108	0.110	0.110	0.161
500	3	-0.017	-2.659	-0.104	0.002	-0.118	-0.471	-	0.120	2.694	0.453	0.282	0.219	0.499	-
	4	0.002	-1.639	-0.014	0.010	-0.165	-0.274	-	0.088	1.659	0.239	0.192	0.225	0.312	-
	6	-0.008	-0.941	-0.033	-0.002	-0.101	-0.126	0.766	0.074	0.954	0.151	0.133	0.162	0.177	0.805
	8	0.001	-0.650	-0.023	0.002	-0.054	-0.064	0.316	0.062	0.660	0.119	0.103	0.114	0.118	0.348
	12	-0.000	-0.406	-0.034	-0.003	-0.026	-0.029	0.107	0.049	0.414	0.090	0.080	0.084	0.084	0.138
1000	3	0.001	-2.642	-0.007	0.010	-0.128	-0.466	-	0.069	2.657	0.270	0.196	0.211	0.479	-
	4	-0.001	-1.647	-0.035	-0.005	-0.177	-0.284	-	0.065	1.656	0.177	0.138	0.209	0.303	-
	6	-0.006	-0.937	-0.024	0.003	-0.098	-0.124	0.761	0.047	0.943	0.109	0.095	0.132	0.152	0.781
	8	0.002	-0.647	-0.020	0.005	-0.052	-0.062	0.312	0.043	0.652	0.084	0.072	0.088	0.093	0.327
	12	-0.001	-0.405	-0.023	-0.001	-0.025	-0.027	0.107	0.034	0.409	0.064	0.054	0.059	0.060	0.123

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.010	-2.643	0.002	0.056	-0.095	-0.453	-	0.107	2.643	0.356	0.296	0.197	0.453	-
	4	0.013	-1.671	-0.091	-0.025	-0.177	-0.277	-	0.091	1.671	0.250	0.175	0.195	0.277	-
	6	-0.007	-0.946	-0.019	0.001	-0.096	-0.128	0.751	0.074	0.946	0.145	0.118	0.134	0.148	0.751
	8	0.002	-0.643	-0.024	0.006	-0.052	-0.060	0.333	0.061	0.643	0.106	0.094	0.094	0.097	0.333
	12	-0.001	-0.401	-0.025	-0.000	-0.022	-0.024	0.112	0.046	0.401	0.078	0.075	0.075	0.077	0.124
500	3	-0.018	-2.679	-0.098	0.024	-0.091	-0.457	-	0.079	2.679	0.315	0.179	0.120	0.457	-
	4	0.000	-1.631	-0.016	0.014	-0.166	-0.271	-	0.059	1.631	0.159	0.130	0.171	0.271	-
	6	-0.007	-0.952	-0.046	-0.008	-0.108	-0.131	0.751	0.048	0.952	0.112	0.088	0.125	0.147	0.751
	8	0.001	-0.650	-0.023	0.004	-0.052	-0.063	0.316	0.042	0.650	0.082	0.066	0.079	0.081	0.316
	12	0.002	-0.404	-0.038	-0.002	-0.026	-0.027	0.104	0.031	0.404	0.065	0.053	0.056	0.059	0.106
1000	3	0.002	-2.633	-0.003	0.021	-0.110	-0.461	-	0.046	2.633	0.179	0.139	0.117	0.461	-
	4	-0.000	-1.649	-0.039	-0.008	-0.180	-0.285	-	0.040	1.649	0.112	0.093	0.180	0.285	-
	6	-0.008	-0.926	-0.022	0.007	-0.094	-0.125	0.758	0.032	0.926	0.072	0.064	0.097	0.125	0.758
	8	0.003	-0.645	-0.025	0.005	-0.050	-0.060	0.313	0.028	0.645	0.059	0.049	0.058	0.065	0.313
	12	-0.000	-0.404	-0.026	-0.001	-0.025	-0.027	0.106	0.022	0.404	0.044	0.036	0.038	0.039	0.106

Table A.6: Simulation results under the design based on Equations (1.23) and (1.24) with $\gamma = 1$, $\beta = 1$

Results for $\hat{\beta}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.008	0.788	0.188	0.020	-0.005	-0.051	-	0.098	0.837	0.521	0.150	0.153	0.095	-
	4	0.001	0.554	0.052	0.009	0.013	0.006	-	0.081	0.588	0.192	0.112	0.085	0.082	-
	6	0.011	0.341	0.052	0.018	0.036	0.032	-0.174	0.074	0.359	0.101	0.075	0.078	0.076	0.244
	8	0.004	0.220	0.017	0.003	0.019	0.017	-0.113	0.049	0.233	0.073	0.057	0.059	0.059	0.144
	12	0.004	0.136	0.010	0.005	0.012	0.011	-0.054	0.040	0.145	0.049	0.043	0.043	0.043	0.075
500	3	0.007	0.781	0.114	0.011	-0.040	-0.051	-	0.064	0.805	0.249	0.100	0.103	0.074	-
	4	0.000	0.540	0.045	0.000	0.008	0.001	-	0.053	0.555	0.128	0.073	0.057	0.056	-
	6	0.006	0.320	0.021	0.005	0.026	0.022	-0.121	0.039	0.329	0.074	0.051	0.055	0.053	0.158
	8	0.007	0.232	0.020	0.013	0.027	0.025	-0.092	0.041	0.239	0.060	0.047	0.052	0.051	0.113
	12	0.001	0.133	0.008	0.002	0.009	0.009	-0.052	0.028	0.137	0.037	0.030	0.031	0.031	0.063
1000	3	0.001	0.762	0.081	0.001	-0.019	-0.054	-	0.050	0.776	0.185	0.080	0.083	0.069	-
	4	-0.002	0.531	0.031	-0.005	0.005	-0.002	-	0.038	0.539	0.091	0.051	0.040	0.040	-
	6	0.003	0.319	0.022	0.004	0.025	0.022	-0.124	0.029	0.323	0.055	0.035	0.041	0.039	0.145
	8	0.002	0.217	0.011	0.002	0.017	0.015	-0.103	0.026	0.221	0.040	0.029	0.034	0.033	0.112
	12	-0.002	0.128	0.002	-0.002	0.005	0.005	-0.053	0.020	0.131	0.027	0.022	0.022	0.022	0.059

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.005	0.763	0.087	0.008	-0.043	-0.059	-	0.063	0.763	0.206	0.091	0.076	0.071	-
	4	-0.007	0.529	0.030	-0.005	0.007	0.000	-	0.060	0.529	0.107	0.075	0.055	0.057	-
	6	0.006	0.339	0.058	0.018	0.028	0.028	-0.171	0.049	0.339	0.079	0.048	0.056	0.055	0.174
	8	0.003	0.217	0.012	-0.002	0.014	0.012	-0.113	0.031	0.217	0.044	0.034	0.037	0.036	0.114
	12	-0.000	0.131	0.008	0.003	0.008	0.007	-0.061	0.026	0.131	0.034	0.031	0.030	0.030	0.061
500	3	0.006	0.762	0.090	0.003	-0.041	-0.054	-	0.043	0.762	0.141	0.069	0.060	0.057	-
	4	-0.001	0.535	0.041	-0.004	0.006	-0.000	-	0.036	0.535	0.080	0.051	0.040	0.039	-
	6	0.007	0.319	0.021	0.000	0.022	0.021	-0.123	0.025	0.319	0.048	0.033	0.036	0.033	0.128
	8	0.003	0.229	0.022	0.010	0.025	0.026	-0.088	0.028	0.229	0.042	0.032	0.034	0.030	0.088
	12	0.002	0.131	0.007	0.001	0.008	0.007	-0.054	0.020	0.131	0.025	0.021	0.022	0.021	0.054
1000	3	0.002	0.751	0.060	-0.007	-0.033	-0.053	-	0.036	0.751	0.102	0.049	0.052	0.054	-
	4	-0.004	0.530	0.025	-0.008	0.004	-0.002	-	0.025	0.530	0.057	0.036	0.029	0.029	-
	6	0.001	0.319	0.017	0.004	0.025	0.022	-0.123	0.019	0.319	0.035	0.024	0.028	0.027	0.123
	8	0.002	0.217	0.010	0.001	0.018	0.015	-0.104	0.019	0.217	0.025	0.019	0.023	0.022	0.104
	12	-0.001	0.128	0.003	-0.002	0.006	0.005	-0.053	0.014	0.128	0.019	0.015	0.015	0.015	0.053

Table A.7: Simulation results under the design based on Equations (1.23) and (1.24) with $\gamma = 1$, $\beta = 1$

Results for $\hat{\gamma}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	-0.003	-2.389	0.066	0.091	-0.406	-0.742	-	0.153	2.450	0.768	0.469	0.488	0.779	-
	4	0.010	-1.449	-0.089	0.043	-0.301	-0.414	-	0.133	1.489	0.375	0.302	0.373	0.466	-
	6	0.013	-0.850	-0.105	0.008	-0.137	-0.179	0.800	0.137	0.888	0.290	0.233	0.252	0.274	0.895
	8	0.005	-0.605	-0.103	0.001	-0.077	-0.097	0.227	0.083	0.626	0.196	0.150	0.164	0.171	0.295
	12	-0.001	-0.361	-0.085	0.018	-0.014	-0.020	0.102	0.069	0.379	0.149	0.112	0.111	0.111	0.169
500	3	0.000	-2.395	-0.077	0.048	-0.438	-0.742	-	0.115	2.430	0.465	0.328	0.496	0.762	-
	4	0.003	-1.451	-0.079	0.029	-0.303	-0.416	-	0.091	1.471	0.285	0.213	0.342	0.444	-
	6	0.010	-0.857	-0.092	0.004	-0.147	-0.186	0.677	0.076	0.872	0.182	0.146	0.198	0.226	0.724
	8	0.006	-0.606	-0.097	0.002	-0.079	-0.097	0.243	0.059	0.615	0.149	0.095	0.122	0.133	0.270
	12	-0.004	-0.387	-0.102	-0.008	-0.039	-0.045	0.070	0.049	0.397	0.133	0.083	0.091	0.094	0.117
1000	3	-0.004	-2.378	-0.084	0.031	-0.458	-0.744	-	0.076	2.393	0.331	0.218	0.494	0.753	-
	4	-0.002	-1.459	-0.076	0.018	-0.311	-0.421	-	0.067	1.470	0.213	0.155	0.332	0.437	-
	6	0.003	-0.846	-0.075	0.012	-0.139	-0.178	0.695	0.049	0.852	0.139	0.095	0.163	0.196	0.715
	8	0.001	-0.601	-0.082	0.005	-0.075	-0.094	0.240	0.044	0.607	0.121	0.080	0.107	0.121	0.261
	12	0.002	-0.378	-0.083	0.001	-0.030	-0.036	0.075	0.034	0.383	0.107	0.058	0.065	0.067	0.100

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.011	-2.376	0.021	0.037	-0.412	-0.729	-	0.103	2.376	0.381	0.307	0.413	0.729	-
	4	0.005	-1.434	-0.101	0.015	-0.299	-0.404	-	0.091	1.434	0.246	0.206	0.300	0.404	-
	6	0.014	-0.848	-0.127	0.015	-0.131	-0.176	0.838	0.081	0.848	0.209	0.185	0.181	0.183	0.838
	8	0.006	-0.618	-0.098	-0.009	-0.092	-0.110	0.214	0.052	0.618	0.133	0.106	0.123	0.129	0.218
	12	-0.005	-0.369	-0.082	0.009	-0.022	-0.029	0.110	0.042	0.369	0.105	0.074	0.079	0.077	0.121
500	3	0.002	-2.388	-0.096	0.039	-0.434	-0.738	-	0.077	2.388	0.300	0.205	0.434	0.738	-
	4	0.002	-1.446	-0.097	0.026	-0.301	-0.411	-	0.062	1.446	0.196	0.142	0.301	0.411	-
	6	0.007	-0.861	-0.094	-0.004	-0.144	-0.189	0.674	0.049	0.861	0.130	0.108	0.150	0.189	0.674
	8	0.008	-0.598	-0.102	0.001	-0.075	-0.096	0.240	0.038	0.598	0.106	0.057	0.080	0.098	0.240
	12	-0.004	-0.385	-0.104	-0.005	-0.037	-0.042	0.072	0.033	0.385	0.105	0.058	0.062	0.062	0.087
1000	3	0.001	-2.370	-0.087	0.050	-0.436	-0.743	-	0.052	2.370	0.201	0.158	0.436	0.743	-
	4	-0.002	-1.458	-0.080	0.013	-0.308	-0.418	-	0.047	1.458	0.138	0.105	0.308	0.418	-
	6	0.002	-0.844	-0.071	0.012	-0.136	-0.176	0.694	0.035	0.844	0.095	0.068	0.137	0.176	0.694
	8	0.000	-0.604	-0.085	0.003	-0.076	-0.099	0.236	0.029	0.604	0.091	0.054	0.082	0.100	0.236
	12	-0.001	-0.379	-0.085	-0.001	-0.030	-0.036	0.078	0.022	0.379	0.086	0.038	0.044	0.047	0.078

Table A.8: Simulation results under the design based on Equations (1.23) and (1.24) with $\gamma = 2$, $\beta = 1$

Results for $\hat{\beta}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.014	0.812	0.678	0.038	-0.006	-0.052	-	0.105	0.875	5.197	0.188	0.130	0.099	-
	4	0.008	0.548	0.088	0.007	0.008	-0.005	-	0.082	0.583	0.209	0.116	0.089	0.083	-
	6	0.004	0.334	0.035	0.011	0.030	0.024	-0.135	0.065	0.355	0.117	0.085	0.079	0.077	0.223
	8	0.025	0.266	0.052	0.043	0.051	0.046	-0.035	0.054	0.273	0.078	0.067	0.067	0.064	0.086
	12	0.000	0.141	-0.002	-0.001	0.011	0.009	-0.047	0.042	0.152	0.058	0.046	0.048	0.047	0.074
500	3	0.012	0.780	0.190	0.022	-0.011	-0.055	-	0.074	0.813	0.370	0.133	0.122	0.083	-
	4	0.003	0.540	0.064	0.001	0.004	-0.006	-	0.057	0.558	0.168	0.084	0.060	0.059	-
	6	0.002	0.325	0.027	0.004	0.025	0.019	-0.115	0.044	0.334	0.083	0.056	0.055	0.052	0.162
	8	0.002	0.230	0.012	0.004	0.024	0.019	-0.085	0.036	0.237	0.063	0.045	0.048	0.046	0.110
	12	0.001	0.141	-0.002	-0.001	0.012	0.010	-0.044	0.030	0.146	0.041	0.033	0.035	0.034	0.059
1000	3	-0.003	0.739	0.116	-0.003	-0.036	-0.063	-	0.048	0.753	0.217	0.081	0.061	0.076	-
	4	0.000	0.528	0.047	-0.005	0.001	-0.009	-	0.038	0.536	0.107	0.054	0.039	0.039	-
	6	-0.000	0.321	0.016	-0.000	0.022	0.016	-0.097	0.032	0.326	0.063	0.042	0.044	0.041	0.126
	8	0.004	0.231	0.011	0.004	0.024	0.019	-0.078	0.028	0.235	0.045	0.033	0.040	0.037	0.093
	12	0.001	0.143	0.001	0.001	0.014	0.011	-0.040	0.022	0.146	0.030	0.025	0.028	0.027	0.050

		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.006	0.778	0.209	0.013	-0.022	-0.052	-	0.065	0.778	0.260	0.115	0.078	0.072	-
	4	0.002	0.523	0.070	-0.004	0.004	-0.007	-	0.055	0.523	0.134	0.077	0.062	0.058	-
	6	0.005	0.330	0.029	0.008	0.029	0.025	-0.118	0.044	0.330	0.070	0.055	0.053	0.051	0.146
	8	0.023	0.268	0.064	0.042	0.053	0.047	-0.034	0.036	0.268	0.064	0.042	0.053	0.047	0.061
	12	-0.000	0.139	-0.003	-0.003	0.010	0.008	-0.045	0.029	0.139	0.039	0.031	0.032	0.030	0.050
500	3	0.012	0.753	0.139	-0.008	-0.031	-0.061	-	0.052	0.753	0.190	0.083	0.056	0.068	-
	4	-0.001	0.524	0.041	-0.007	0.001	-0.009	-	0.037	0.524	0.095	0.056	0.041	0.042	-
	6	0.000	0.323	0.023	0.002	0.025	0.018	-0.113	0.031	0.323	0.053	0.037	0.035	0.033	0.119
	8	0.004	0.229	0.014	0.004	0.021	0.018	-0.084	0.023	0.229	0.042	0.030	0.031	0.030	0.086
	12	0.000	0.137	-0.003	-0.003	0.009	0.007	-0.047	0.020	0.137	0.028	0.022	0.021	0.021	0.048
1000	3	-0.003	0.741	0.108	-0.003	-0.036	-0.065	-	0.035	0.741	0.128	0.057	0.043	0.066	-
	4	-0.002	0.527	0.045	-0.005	0.002	-0.009	-	0.026	0.527	0.066	0.035	0.025	0.026	-
	6	-0.001	0.316	0.014	-0.003	0.021	0.014	-0.096	0.022	0.316	0.043	0.026	0.028	0.025	0.099
	8	0.004	0.231	0.010	0.004	0.024	0.019	-0.078	0.018	0.231	0.028	0.021	0.027	0.026	0.078
	12	0.001	0.142	0.000	0.001	0.013	0.011	-0.042	0.015	0.142	0.020	0.017	0.019	0.018	0.042

Table A.9: Simulation results under the design
based on Equations (1.23) and (1.24) with $\gamma = 2$, $\beta = 1$

Results for $\hat{\gamma}$

		MEAN BIAS							RMSE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.016	-2.012	0.464	0.213	-0.836	-1.088	-	0.196	2.099	4.998	0.639	0.903	1.118	-
	4	0.018	-1.241	-0.093	0.081	-0.504	-0.632	-	0.166	1.307	0.530	0.390	0.570	0.681	-
	6	0.003	-0.759	-0.174	0.020	-0.227	-0.291	0.610	0.124	0.799	0.350	0.232	0.299	0.348	0.723
	8	0.065	-0.514	-0.098	0.091	-0.080	-0.117	0.296	0.145	0.553	0.197	0.230	0.189	0.212	0.408
	12	0.002	-0.372	-0.205	-0.006	-0.054	-0.071	0.043	0.091	0.399	0.261	0.141	0.148	0.153	0.166
500	3	0.004	-1.995	0.007	0.209	-0.778	-1.078	-	0.154	2.047	0.742	0.507	0.821	1.097	-
	4	0.010	-1.255	-0.151	0.062	-0.512	-0.639	-	0.115	1.286	0.375	0.282	0.545	0.662	-
	6	-0.000	-0.766	-0.160	0.016	-0.234	-0.297	0.610	0.087	0.785	0.261	0.161	0.272	0.325	0.664
	8	0.000	-0.562	-0.176	0.004	-0.127	-0.167	0.166	0.072	0.577	0.236	0.128	0.173	0.202	0.228
	12	0.001	-0.369	-0.184	-0.007	-0.052	-0.070	0.039	0.058	0.381	0.214	0.093	0.105	0.113	0.116
1000	3	0.005	-2.050	-0.108	0.139	-0.878	-1.104	-	0.093	2.071	0.441	0.301	0.912	1.112	-
	4	0.008	-1.261	-0.143	0.048	-0.519	-0.645	-	0.080	1.273	0.291	0.179	0.535	0.655	-
	6	0.001	-0.763	-0.165	0.011	-0.231	-0.295	0.616	0.061	0.773	0.219	0.120	0.252	0.310	0.642
	8	0.005	-0.554	-0.159	0.008	-0.120	-0.160	0.183	0.054	0.562	0.198	0.091	0.146	0.179	0.216
	12	0.002	-0.366	-0.158	-0.005	-0.050	-0.068	0.042	0.042	0.373	0.176	0.065	0.081	0.093	0.085
		MEDIAN BIAS							MAE						
n	T	INF	ML	HK	PCML	MPL	MML	SPJ	INF	ML	HK	PCML	MPL	MML	SPJ
250	3	0.002	-2.019	-0.080	0.145	-0.838	-1.102	-	0.135	2.019	0.644	0.388	0.838	1.102	-
	4	0.008	-1.254	-0.147	0.062	-0.506	-0.636	-	0.119	1.254	0.389	0.276	0.506	0.636	-
	6	0.006	-0.756	-0.190	0.031	-0.224	-0.290	0.603	0.084	0.756	0.264	0.156	0.226	0.290	0.604
	8	0.061	-0.576	-0.088	0.056	-0.119	-0.180	0.276	0.069	0.576	0.113	0.131	0.132	0.189	0.276
	12	0.002	-0.375	-0.223	-0.015	-0.055	-0.076	0.037	0.060	0.375	0.227	0.089	0.099	0.108	0.111
500	3	-0.017	-1.995	-0.137	0.185	-0.780	-1.072	-	0.089	1.995	0.492	0.362	0.780	1.072	-
	4	0.011	-1.260	-0.161	0.057	-0.515	-0.638	-	0.080	1.260	0.262	0.192	0.515	0.638	-
	6	-0.003	-0.768	-0.170	0.010	-0.232	-0.298	0.609	0.058	0.768	0.197	0.107	0.232	0.298	0.609
	8	-0.002	-0.562	-0.176	0.003	-0.126	-0.165	0.168	0.050	0.562	0.183	0.089	0.129	0.166	0.175
	12	0.001	-0.369	-0.188	-0.010	-0.053	-0.069	0.042	0.038	0.369	0.189	0.064	0.074	0.079	0.082
1000	3	-0.004	-2.024	-0.131	0.129	-0.850	-1.102	-	0.064	2.024	0.312	0.205	0.850	1.102	-
	4	0.008	-1.256	-0.148	0.055	-0.516	-0.645	-	0.053	1.256	0.200	0.108	0.516	0.645	-
	6	0.003	-0.766	-0.163	0.012	-0.233	-0.294	0.616	0.041	0.766	0.167	0.077	0.233	0.294	0.616
	8	0.004	-0.552	-0.164	0.004	-0.118	-0.158	0.183	0.037	0.552	0.165	0.063	0.118	0.158	0.183
	12	0.002	-0.363	-0.157	-0.002	-0.046	-0.065	0.041	0.028	0.363	0.157	0.046	0.055	0.068	0.056

Appendix B

Approximation of $p(y_{it}|s_{it})$

This appendix includes the derivation of the result reported in Equations (2.6) and (2.7). The joint probability for the outcome variables is given by

$$
\begin{aligned}
&p(y_{it}=1, s_{it}=1|\alpha_i,\eta_i,\boldsymbol{x}_{it},\boldsymbol{w}_{it})) = \\
&\qquad \int_{u_{it}>-\bar{u}_{it}}\int_{\varepsilon_{it}>-\bar{\varepsilon}_{it}} f(\varepsilon_{it}|u_{it})f(u_{it})\mathrm{d}\varepsilon_{it}\mathrm{d}u_{it} = \\
&\qquad\qquad \int_{u_{it}>-\bar{u}_{it}} \Psi\left(\frac{\alpha_i+\boldsymbol{x}_{it}'\boldsymbol{\beta}+\delta+u_{it}\rho}{\sqrt{1-\rho^2}}\right)\psi(u_{it})\mathrm{d}u_{it},
\end{aligned}
$$

where $\Psi(\cdot)$ denotes the the standard logistic distribution function, so that $\Psi(w)=\frac{\exp(w)}{1+\exp(w)}$, $\psi(\cdot)$ is the relative density function, that is $\psi(w)=\Psi(w)[1-\Psi(w)]$, and $\bar{\varepsilon}_{it}$ and $\bar{u}_{it}$ are the linear indeces $\bar{\varepsilon}_{it}=\alpha_i+\boldsymbol{x}_{it}'\boldsymbol{\beta}+\delta s_{it}$ and $\bar{u}_{it}=\eta_i+\boldsymbol{w}_{it}'\boldsymbol{\gamma}$.

Similarly, we have

$$
\begin{aligned}
&p(y_{it}=0, s_{it}=1|\alpha_i,\eta_i,\boldsymbol{x}_{it},\boldsymbol{w}_{it})) = \\
&\qquad \int_{u_{it}>-\bar{u}_{it}}\left[1-\Psi\left(\frac{\alpha_i+\boldsymbol{x}_{it}'\boldsymbol{\beta}+\delta+u_{it}\rho}{\sqrt{1-\rho^2}}\right)\right]\psi(u_{it})\mathrm{d}u_{it},
\end{aligned}
$$

$$p(y_{it}=1,s_{it}=0|\alpha_i,\eta_i,\boldsymbol{x}_{it},\boldsymbol{w}_{it}))=$$
$$\int_{u_{it}<-\bar{u}_{it}}\left[\Psi\left(\frac{\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+u_{it}\rho}{\sqrt{1-\rho^2}}\right)\right]\psi(u_{it})\mathrm{d}u_{it},$$

and

$$p(y_{it}=0,s_{it}=0|\alpha_i,\eta_i,\boldsymbol{x}_{it},\boldsymbol{w}_{it}))=$$
$$\int_{u_{it}<-\bar{u}_{it}}\left[1-\Psi\left(\frac{\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+u_{it}\rho}{\sqrt{1-\rho^2}}\right)\right]\psi(u_{it})\mathrm{d}u_{it}.$$

The joint probability of $(y_{it}=1,s_{it}=1)$ evaluated at $\rho=0$ is

$$p_0(y_{it}=1,s_{it}=1)=\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma}),$$

and its derivative with respect to ρ is

$$\frac{\partial p_0(y_{it}=1,s_{it}=1)}{\partial\rho}=$$
$$\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\beta+\delta)]\int_{u_{it}>-\bar{u}_{it}}u_{it}\psi(u_{it})du_{it}.$$

Summing the previous elements, we derive an approximated joint log-probability of Equation (2.6), so we have

$$\log p_\rho(y_{it}=1,s_{it}=1)\approx\log\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)$$
$$+\log\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})+\frac{\rho[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)]}{\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})}\int_{u_{it}>-\bar{u}_{it}}u_{it}\psi(u_{it})du_{it}.$$

Then, conditioning the latter on the probability of $s_{it}=1$, we end up with the following conditional log-probability

$$\log p_\rho(y_{it}=1|s_{it}=1)\approx\log\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)+$$
$$\frac{\rho[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)]}{\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})}\int_{u_{it}>-\bar{u}_{it}}u_{it}\psi(u_{it})du_{it}.$$

Using a similar argument, we have

$$p_0(y_{it}=0, s_{it}=1) = [1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)]\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})$$

and

$$\frac{\partial p_0(y_{it}=0, s_{it}=1)}{\partial\rho} = \\ -\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)]\int_{u_{it}>-\bar{u}_{it}} u_{it}\psi(u_{it})du_{it}.$$

Then, we have

$$\log p_\rho(y_{it}=0, s_{it}=1) \approx \log[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)] + \\ \log\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma}) - \frac{\rho\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)}{\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})}\int_{u_{it}>-\bar{u}_{it}} u_{it}\psi(u_{it})du_{it},$$

so that

$$\log p_\rho(y_{it}=0|s_{it}=1) \approx \log[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)] - \\ \frac{\delta\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\delta)}{\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})}\int_{u_{it}>-\bar{u}_{it}} u_{it}\psi(u_{it})du_{it}.$$

Finally, computing the log-odds ratio for the approximated conditional probabilities, we have the linear index in Equation (B.1), that is

$$\log\frac{p_\rho(y_{it}=1|s_{it}=1)}{p_\rho(y_{it}=0|s_{it}=1)} \approx \alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta} + \\ \frac{\rho}{\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})}\int_{u_{it}>-\bar{u}_{it}} u_{it}\psi(u_{it})du_{it} = \\ \alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\rho\,\mathrm{E}(u_{it}|u_{it}>-\bar{u}_{it}). \tag{B.1}$$

This result allows us to recognise the approximation in Equation (2.7) as a logit-type probability.

The same methodology can be applied for $(y_{it}=1, s_{it}=0)$, so that the joint probability evaluated at $\rho = 0$ is

$$p_0(y_{it} = 1, s_{it} = 0) = \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})[1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})],$$

and

$$\frac{\partial p_0(y_{it} = 1, s_{it} = 0)}{\partial \rho} = \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})[1 - \Psi(\alpha_i + \boldsymbol{x}'_{it}\beta)] \int_{u_{it} < -\bar{u}_{it}} u_{it}\psi(u_{it}) du_{it}.$$

The joint probability becomes

$$\log p_\rho(y_{it} = 1, s_{it} = 0) \approx \log \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}) + \log[1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})] + \frac{\rho[1 - \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]}{[1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]} \int_{u_{it} < -\bar{u}_{it}} u_{it}\psi(u_{it}) du_{it}.$$

Again, conditioning on the probability of $s_{it} = 0$, we end up with the following conditional log-probability

$$\log p_\rho(y_{it} = 1 | s_{it} = 0) \approx \log \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}) + \frac{\rho[1 - \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]}{[1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]} \int_{u_{it} < -\bar{u}_{it}} u_{it}\psi(u_{it}) du_{it}.$$

Using a similar argument, we have

$$p_0(y_{it} = 0, s_{it} = 0) = [1 - \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})][1 - \Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]$$

and

$$\frac{\partial p_0(y_{it} = 0, s_{it} = 0)}{\partial \rho} = -\Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})[1 - \Psi(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})] \int_{u_{it} < -\bar{u}_{it}} u_{it}\psi(u_{it}) du_{it}.$$

Then, we get

$$\log p_\rho(y_{it}=0,s_{it}=0) \approx \log[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta})] + \\ \log[1-\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})] - \frac{\rho\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta})}{[1-\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})]}\int_{u_{it}<-\bar{u}_{it}} u_{it}\psi(u_{it})du_{it},$$

so that

$$\log p_\rho(y_{it}=0|s_{it}=1) \approx \log[1-\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta})] - \\ \frac{\delta\Psi(\alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta})}{[1-\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})]}\int_{u_{it}<-\bar{u}_{it}} u_{it}\psi(u_{it})du_{it}.$$

Finally, the log-odds ratio equals again

$$\log\frac{p_\rho(y_{it}=1|s_{it}=0)}{p_\rho(y_{it}=0|s_{it}=0)} \approx \alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta} + \\ \frac{\rho}{[1-\Psi(\eta_i+\boldsymbol{w}'_{it}\boldsymbol{\gamma})]}\int_{u_{it}<-\bar{u}_{it}} u_{it}\psi(u_{it})du_{it} = \\ \alpha_i+\boldsymbol{x}'_{it}\boldsymbol{\beta}+\rho\,\mathrm{E}(u_{it}|u_{it}<-\bar{u}_{it}). \tag{B.2}$$

The proofs for the sample selection model and the dynamic models are straightforward.

Appendix C

Derivation of the $E[x|x < a]$

What follows is relative to the result reported in Equation (2.10). The conditional expected value of a random variable x is defined as

$$\mathrm{E}[x|x < a] = \int_{-\infty}^{a} x f(x|x < a) dx$$

with

$$f(x|x < a) = \frac{f(x)}{[Prob(x < a)]} = \frac{f(x)}{[F(a)]}$$

where $f(x)$ and $F(x)$ are the density function and the distribution function of x, respectively.

From the logistic distribution we have

$$F(x) = \Psi(x) = \frac{\exp(x)}{1 + \exp(x)}$$

and

$$f(x) = \psi(x) = \frac{\exp(x)}{(1 + \exp(x))^2}$$

We are interested in $\mathrm{E}[x|x < a]$, so we define

$$\mathrm{E}[x|x < a] = \frac{1}{\Psi(a)} \int_{-\infty}^{a} x \frac{\exp(x)}{(1 + \exp(x))^2} dx$$

that can be computed by

$$\frac{1}{\Psi(a)}\left[\left(\frac{x\exp(x)}{1+\exp(x)} - \log(1+\exp(x))\right)\Bigg|_{-\infty}^{a}\right].$$

The solution of the definite integral is

$$\frac{1}{\Psi(a)}\left(\frac{a\exp(a)}{1+\exp(a)} - \log(1+\exp(a))\right).$$

Since $\Psi(a) = \frac{\exp(a)}{1+\exp(a)}$ we multiply the terms and we get

$$\frac{a\exp(a)}{1+\exp(a)} \times \frac{1+\exp(a)}{\exp(a)} - \frac{\log(1+\exp(a))}{\Psi(a)}.$$

The previous expression can be written by

$$\mathrm{E}[x|x < a] = a + \frac{log(1-\Psi(a))}{\Psi(a)}.$$

In our case, $a = -(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})$. We exploit the symmetry of the logistic distribution function and we have

$$\mathrm{E}(u_{it}|u_{it} < -\bar{u}_{it}) = -(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma}) + \frac{\ln[\Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]}{[1-\Psi(\eta_i + \boldsymbol{w}'_{it}\boldsymbol{\gamma})]}.$$

Appendix D

Recursions for the Score and the Hessian Matrix

Static logit model

Define now the the first and second derivatives with respect to the arguments of the function, the vector $\boldsymbol{\phi}$. For $h, j = 1, \ldots, T$, $s = 0, \ldots, t$, and $t = 1, \ldots, T$, we have

$$
\begin{aligned}
f_{t,s}^{(h)}(\boldsymbol{\phi}) &= \frac{\partial f_{t,s}(\boldsymbol{\phi})}{\partial \phi_h}, \\
f_{t,s}^{(h,j)}(\boldsymbol{\phi}) &= \frac{\partial^2 f_{t,s}(\boldsymbol{\phi})}{\partial \phi_h \partial \phi_j}.
\end{aligned}
$$

For what concerns the computations of the derivatives we can exploit the same recursive structure presented in the Equation (3.6), so that

1. for $t = 1$ the first derivatives are

$$
\begin{aligned}
f_{1,0}^{(h)}(\boldsymbol{\phi}) &= 0, \quad h = 1, \ldots, T, \\
f_{1,1}^{(h)}(\boldsymbol{\phi}) &= \begin{cases} \exp(\phi_1), & h = 1, \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}
$$

and the second derivatives are

$$\begin{aligned}
f_{1,0}^{(h,j)}(\boldsymbol{\phi}) &= 0, \quad h,j=1,\ldots,T,\\
f_{1,1}^{(j,j)}(\boldsymbol{\phi}) &= \begin{cases} \exp(\phi_1), & h=j=1,\\ 0, & \text{otherwise;} \end{cases}
\end{aligned}$$

2. for $t=2,\ldots,T$ and $s=1,\ldots,t$, we compute the first derivatives as

$$\begin{aligned}
f_{t,0}^{(h)}(\boldsymbol{\phi}) &= 0, \quad h=1,\ldots,T,\\
f_{t,s}^{(h)}(\boldsymbol{\phi}) &= \begin{cases} f_{t-1,s}^{(h)}(\boldsymbol{\phi}) + f_{t-1,s-1}^{(h)}(\boldsymbol{\phi})\exp(\phi_t), & h=1,\ldots,t-1,\\ f_{t-1,s-1}(\boldsymbol{\phi})\exp(\phi_t), & h=t,\\ 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

and second derivatives as follwing

$$\begin{aligned}
f_{t,0}^{(h,j)}(\boldsymbol{\phi}) &= 0, \quad h,j=1,\ldots,T,\\
f_{t,s}^{(h,j)}(\boldsymbol{\phi}) &= \begin{cases} f_{t-1,s}^{(h,j)}(\boldsymbol{\phi}) + f_{t-1,s-1}^{(h,j)}(\boldsymbol{\phi})\exp(\phi_t), & h,j=1,\ldots,t-1,\\ f_{t-1,s-1}^{(h)}(\boldsymbol{\phi})\exp(\phi_t), & h=1,\ldots,t-1,\ j=t\\ f_{t-1,s-1}^{(j)}(\boldsymbol{\phi})\exp(\phi_t), & h=t,\ j=1,\ldots,t-1,\\ f_{t-1,s-1}(\boldsymbol{\phi})\exp(\phi_t), & h=j=t,\\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

Simplified QE

As for the static model, define now the first and the second derivatives

$$\begin{aligned}
g_{t,s}^{(h)}(\boldsymbol{\phi}) &= \frac{\partial f_{t,s}(\boldsymbol{\phi})}{\partial \phi_h},\\
g_{t,s}^{(h,j)}(\boldsymbol{\phi}) &= \frac{\partial^2 f_{t,s}(\boldsymbol{\phi})}{\partial \phi_h \partial \phi_j},
\end{aligned}$$

respectively, where these quantities are computed for $h,j=1,\ldots,T$, $s=0,\ldots,t$, and $t=1,\ldots,T$ and where we further define $\phi_{T+1}\equiv\psi$ in order to include the derivative with respect to the state dependence parameter which is an additional argument of our function further to the T elements of $\boldsymbol{\phi}$.

Regarding the first derivatives of the function in Equation (3.9), we exploit the same recursion:

1. for $t = 1$ compute

$$g^{(h)}_{1,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} \exp(\phi_1 + a\psi), & h = 1, \\ a\exp(\phi_1 + a\psi), & h = T + 1, \end{cases}$$

   for $a = 0, 1$ and $s = v = 1$ and $g^{(h)}_{1,a,s,v} = 0$ in all other cases;

2. for $t = 2, \ldots, T$ consider the following cases:

   - for $s = 1, \ldots, t-1$ and $v = 0$,

   $$g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = g^{(h)}_{t-1,a,s,0}(\boldsymbol{\phi}, \psi) + g^{(h)}_{t-1,a,s,1}(\boldsymbol{\phi}, \psi), \quad h = 1, \ldots, t-1, T+1;$$

   - for $s = 1, \ldots, t$, $v = 1$,

   $$g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi), & h = 1, \ldots, t-1, \\ g_{t,a,s,v}(\boldsymbol{\phi}, \psi), & h = t, \\ g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi) & \\ +g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi), & h = T + 1; \end{cases}$$

   - $g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = 0$ in all other cases.

Regarding the second derivatives, we have:

1. for $t = 1$ compute

$$g^{(h,j)}_{1,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} \exp(\phi_1 + a\psi), & h = j = 1, \\ a\exp(\phi_1 + a\psi), & h = 1,\ j = T + 1, \\ a\exp(\phi_1 + a\psi), & h = T + 1,\ j = 1, \\ a\exp(\phi_1 + a\psi), & h, j = T + 1, \end{cases}$$

   for $a = 0, 1$ and $s = v = 1$ and $g^{(h)}_{1,a,s,v} = 0$ in all other cases.

2. for $t = 2, \ldots, T$ consider the following cases:

- for $s = 1, \dots, t-1$ and $v = 0$,

$$g^{(h,j)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = g^{(h,j)}_{t-1,a,s,0}(\boldsymbol{\phi}, \psi) + g^{(h,j)}_{t-1,a,s,1}(\boldsymbol{\phi}, \psi), \quad h, j = 1, \dots, t-1, T+1;$$

- for $s = 1, \dots, t$, $v = 1$,

$$g^{(h,j)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi), & h, j = 1, \dots, t-1, \\ g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi), & h = 1, \dots, t-1,\ j = t, \\ g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi), & \\ +g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi) & h = 1, \dots, t-1,\ j = T+1, \\ g^{(j)}_{t,a,s,v}(\boldsymbol{\phi}, \psi), & h = t,\ j = 1, \dots, t, T+1, \\ g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi) & \\ +g^{(j)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi), & h = T+1,\ j = 1, \dots, t-1 \\ g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi), & h = T+1,\ j = t, \\ g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi) & \\ +g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi) + g^{(j)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi) & \\ +g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t+\psi), & h = T+1,\ j = T+1; \end{cases}$$

- $g^{(h,j)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = 0$ in all other cases.

Pseudo Conditional Maximum Likelihood estimator

For what concerns the first derivatives of the function in Equation (3.10), compute:

1. for $t = 1$,

   - for $s = v = 0$ and $a = 0, 1$,

   $$g^{(h)}_{1,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} a\exp(a\nu_1)(-q_{i1}), & h = T+1, \\ 0, & \text{otherwise}; \end{cases}$$

   - for for $s = v = 1$ and $a = 0, 1$,

   $$g^{(h)}_{1,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} \exp(\phi_1 + a(\nu_1+\gamma)), & h = 1, \\ a\exp(\phi_1 + a(\nu_1+\gamma))(1-q_{i1}), & h = T+1; \end{cases}$$

- $g^{(h)}_{1,a,s,v} = 0$ in all other cases.

2. for $t = 2, \dots, T$ consider the following cases:

   - for $s = 1, \dots, t-1$ and $v = 0$,

$$g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} g^{(h)}_{t-1,a,s,0}(\boldsymbol{\phi}, \gamma) + g^{(h)}_{t-1,a,s,1}(\boldsymbol{\phi}, \gamma)\exp(\nu_t), & h = 1, \dots, t, \\ \\ g^{(h)}_{t-1,a,s,0}(\boldsymbol{\phi}, \gamma) + g^{(h)}_{t-1,a,s,1}(\boldsymbol{\phi}, \gamma)\exp(\nu_t) & \\ +g_{t-1,a,s,1}(\boldsymbol{\phi}, \gamma)\exp(\nu_t)(-q_{it}), & h = T+1; \end{cases}$$

   - for $s = 1, \dots, t$, $v = 1$,

$$g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \gamma)\exp(\phi_t) + g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \gamma)\exp(\phi_t + \nu_t + \gamma) + & \\ g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \gamma)\exp(\phi_t) + g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \gamma)\exp(\phi_t + \nu_t + \gamma) & h = 1, \dots, t \\ \\ g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \gamma)\exp(\phi_t) + g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \gamma)\exp(\phi_t + \nu_t + \gamma) & \\ +g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \gamma)\exp(\phi_t + \nu_t + \gamma)(1 - q_{it}), & h = T+1; \end{cases}$$

   - $g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \gamma) = 0$ in all other cases.

Following the same approach as above, the second derivatives are:

1. for $t = 1$,

   - with $v = 0$ compute

$$g^{(h,j)}_{1,a,s,0}(\boldsymbol{\phi}, \psi) = \ a\exp(a\nu_1)(-q_{i1})^2, \quad h, j = T+1;$$

   - while for $v = 1$ we have

$$g^{(h,j)}_{1,a,s,1}(\boldsymbol{\phi}, \psi) = \begin{cases} \exp(\phi_1 + a(\nu_1 + \gamma)), & h = j = 1, \\ a\exp(\phi_1 + a(\nu_1 + \gamma))(1 - q_{i1}), & h = 1,\ j = T+1, \\ a\exp(\phi_1 + a(\nu_1 + \gamma))(1 - q_{i1}), & h = T+1,\ j = 1, \\ a\exp(\phi_1 + a(\nu_1 + \gamma))(1 - q_{i1})^2, & h, j = T+1; \end{cases}$$

   - $g^{(h,j)}_{1,a,s,v}(\boldsymbol{\phi}, \psi) = 0$ in other cases.

2. for $t = 2, \dots, T$:

   - for $s = 1 \dots t-1$ and $v = 0$, we have:

$$g^{(h,j)}_{t,a,s,v}(\boldsymbol{\phi},\gamma) = \begin{cases} g^{(h,j)}_{t-1,a,s,0}(\boldsymbol{\phi},\gamma) + g^{(h,j)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t), & h,j = 1,\dots,t \\ g^{(h,j)}_{t-1,a,s,0}(\boldsymbol{\phi},\gamma) + g^{(h,j)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t), \\ +g^{(j)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t)(-q_{it}) & h = T+1,\ j = 1,\dots,t \\ g^{(h,j)}_{t-1,a,s,0}(\boldsymbol{\phi},\gamma) + g^{(h,j)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t), \\ +g^{(h)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t)(-q_{it}), & h = 1,\dots,t,\ j = T+1, \\ g^{(h,j)}_{t-1,a,s,0}(\boldsymbol{\phi},\gamma) + g^{(h,j)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t), \\ +g^{(h)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t)(-q_{it}) + g^{(j)}_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\nu_t)(-q_{it}) \\ +g_{t-1,a,s,1}\exp(\nu_t)(-q_{it})^2 & h = T+1,\ j = T+1; \end{cases}$$

   - for $s = 1 \dots t-1$ and $v = 1$, we have:

$$g^{(h,j)}_{t,a,s,v}(\boldsymbol{\phi},\gamma) = \begin{cases} g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)+ \\ +g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)+ \\ +g^{(j)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)+ \\ g_{t-1,a,s,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma) & h,j = 1,\dots,t, \\ g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)+ \\ g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)(1-q_{it})+ \\ g^{(j)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma) \\ +g_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)(1-q_{it}) & h = 1,\dots,t,\ j = T+1, \\ g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)+ \\ g^{(j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)(1-q_{it})+ \\ g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma) \\ +g_{t-1,a,s,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)(1-q_{it}) & h = 1,\dots,t,\ j = T+1, \\ g^{(h,j)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t) + g^{(h,j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)+ \\ g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)(1-q_{it})+ \\ g^{(j)}_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)(1-q_{it})+ \\ g_{t-1,a,s-1,1}(\boldsymbol{\phi},\gamma)\exp(\phi_t+\nu_t+\gamma)(1-q_{it})^2, & h = T+1,\ j = T+1; \end{cases}$$

3. $g^{(h,j)}_{t,a,s,v}(\boldsymbol{\phi},\gamma) = 0$ in all other cases.

Modified Q.E.

Regarding the first derivatives of the function reported in Equation (3.11), we exploit the same recursion:

1. with $t = 1$

   - for $s = v = 0$ compute

   $$g^{(h)}_{1,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} 0, & h = 1, \\ \mathbb{1}\{a = 0\} \exp(\phi_1 + \mathbb{1}\{a = 0\}\psi), & h = T + 1; \end{cases}$$

   - for $s = v = 1$ compute

   $$g^{(h)}_{1,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} \exp(\phi_1 + \mathbb{1}\{a = 1\}\psi), & h = 1, \\ \mathbb{1}\{a = 1\} \exp(\phi_1 + \mathbb{1}\{a = 1\}\psi), & h = T + 1; \end{cases}$$

   - $g^{(h)}_{1,a,s,v} = 0$ in all other cases.

2. with $t = 2, \dots, T$ consider the following cases:

   - for $s = 0, \dots, t - 1$ and $v = 0$,

   $$g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} 0, & h = 1, \dots, t - 1 \\ g^{(h)}_{t-1,a,s,0}(\boldsymbol{\phi}, \psi) \exp(\psi) + g_{t-1,a,s,0} \exp(\psi) + g^{(h)}_{t-1,a,s,1}(\boldsymbol{\phi}, \psi), & h = T + 1; \end{cases}$$

   - for $s = 1, \dots, t$, $v = 1$,

   $$g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi) \exp(\phi_t) + g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi) \exp(\phi_t + \psi) & \\ +g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi) \exp(\phi_t) + +g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi) \exp(\phi_t + \psi), & h = 1, \dots, t, \\ g^{(h)}_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi) \exp(\phi_t) + g^{(h)}_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi) \exp(\phi_t + \psi) & \\ +g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi) \exp(\phi_t + \psi), & h = T + 1; \end{cases}$$

3. $g^{(h)}_{t,a,s,v}(\boldsymbol{\phi}, \psi) = 0$ in all other cases.

Regarding the second derivatives, we have:

1. for $t = 1$ and $v = 0$ compute

   $$g^{(h,j)}_{1,a,s,v}(\boldsymbol{\phi}, \psi) = \mathbb{1}\{a = 0\} \exp(\mathbb{1}\{a = 0\}\psi), \quad h, j = T + 1;$$

2. for $t = 1$ and $v = 1$ compute

$$g_{1,a,s,v}^{(h,j)}(\boldsymbol{\phi}, \psi) = \begin{cases} \exp(\phi_1 + \mathbb{1}\{a=1\}\psi), & h = j = 1, \\ \mathbb{1}\{a=1\}\exp(\phi_1 + \mathbb{1}\{a=1\}\psi), & h = 1,\ j = T+1, \\ \mathbb{1}\{a=1\}\exp(\phi_1 + \mathbb{1}\{a=1\}\psi), & h = T+1,\ j = 1, \\ \mathbb{1}\{a=1\}\exp(\phi_1 + \mathbb{1}\{a=1\}\psi), & h, j = T+1; \end{cases}$$

   for $a = 0, 1$ and $s = v = 1$ and $g_{1,a,s,v}^{(h)} = 0$ in all other cases.

3. for $t = 2, \dots, T$ consider the following cases:

   - for $s = 1, \dots, t-1$ and $v = 0$,

$$g_{t,a,s,v}^{(h,j)}(\boldsymbol{\phi}, \psi) = \begin{cases} g_{t-1,a,s,0}^{(h,j)}(\boldsymbol{\phi}, \psi) + \exp(\psi) + g_{t-1,a,s,1}^{(h,j)}(\boldsymbol{\phi}, \psi) \quad h = j = 1, \dots, t-1; \\ g_{t-1,a,s,0}^{(h,j)}(\boldsymbol{\phi}, \psi) + \exp(\psi) + g_{t-1,a,s,1}^{(h,j)}(\boldsymbol{\phi}, \psi) \\ +g_{t-1,a,s,0}^{(h)}(\boldsymbol{\phi}, \psi)\exp(\psi), \quad h = 1, \dots, t-1, j = T+1; \\ g_{t-1,a,s,0}^{(h,j)}(\boldsymbol{\phi}, \psi) + \exp(\psi) + g_{t-1,a,s,1}^{(h,j)}(\boldsymbol{\phi}, \psi) \\ +g_{t-1,a,s,0}^{(j)}(\boldsymbol{\phi}, \psi)\exp(\psi), \quad j = 1, \dots, t-1, h = T+1; \\ g_{t-1,a,s,0}^{(h,j)}(\boldsymbol{\phi}, \psi) + \exp(\psi) + g_{t-1,a,s,1}^{(h,j)}(\boldsymbol{\phi}, \psi) \\ +g_{t-1,a,s,0}^{(h)}(\boldsymbol{\phi}, \psi)\exp(\psi) + g_{t-1,a,s,0}^{(j)}(\boldsymbol{\phi}, \psi)\exp(\psi) \\ +g_{t-1,a,s,0}(\boldsymbol{\phi}, \psi)\exp(\psi), \quad j = h = T+1. \end{cases}$$

   - for $s = 1, \dots, t$, $v = 1$,

$$g_{t,a,s,v}^{(h,j)}(\boldsymbol{\phi}, \psi) = \begin{cases} g_{t-1,a,s-1,0}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi), & h, j = 1, \dots, t-1, \\ g_{t,a,s,v}^{(h)}(\boldsymbol{\phi}, \psi), & h = 1, \dots, t-1,\ j = t, \\ g_{t-1,a,s-1,0}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi), & \\ +g_{t-1,a,s-1,1}^{(h)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi) & h = 1, \dots, t-1,\ j = T+1, \\ g_{t,a,s,v}^{(j)}(\boldsymbol{\phi}, \psi), & h = t,\ j = 1, \dots, t, T+1, \\ g_{t-1,a,s-1,0}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi) & \\ +g_{t-1,a,s-1,1}^{(j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi), & h = T+1,\ j = 1, \dots, t-1 \\ g_{t,a,s,v}^{(h)}(\boldsymbol{\phi}, \psi), & h = T+1,\ j = t, \\ g_{t-1,a,s-1,0}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi) & \\ +g_{t-1,a,s-1,1}^{(h)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi) + g_{t-1,a,s-1,1}^{(j)}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi) & \\ +g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi)\exp(\phi_t + \psi), & h = T+1,\ j = T+1; \end{cases}$$

   - $g_{t,a,s,v}^{(h,j)}(\boldsymbol{\phi}, \psi) = 0$ in all other cases.