Conditional inference for binary panel data models with predetermined covariates

Claudia Pigini[1]

Marche Polytechnic University, Ancona, Italy

Francesco Bartolucci

University of Perugia, Italy

Abstract

A fixed-effects logit model that accounts for feedback effects of the dependent variable on the covariates is proposed. The model is formulated by including leads of the predetermined covariates among the regressors and it is proved to satisfy certain theoretical properties under some regularity conditions on the distribution of the covariates. Estimation is based on the Conditional Maximum Likelihood (CML) method for the static logit model and the Pseudo-CML (PCML) method for the corresponding dynamic formulation. Both methods have good finite-sample properties even when the required regularity conditions are not satisfied. An application is provided about female labor supply where we jointly account for the predetermined number of children and husbands' income. Differently from previous studies, it emerges that female employment history does not affect future fertility choices and the husband's earnings, as no evidence of feedback effects is found.

Keywords: Binary panel data, feedback effects, fixed effects, conditional maximum likelihood, female labor supply

[1]Corresponding author: Claudia Pigini, Department of Economics and Social Sciences, Marche Polytechnic University, P.le Martelli, 8, 60121 Ancona (IT), *email*: c.pigini@univpm.it

1. Introduction

A wide range of empirical microeconomic applications requires the estimation of binary, and possibly dynamic, panel data models. One popular approach relies on conditional inference methods for logit models for binary panel data (Andersen, 1970; Chamberlain, 1980). In particular, Conditional Maximum Likelihood (CML) may be applied to estimate the fixed-effects logit model because this model admits sufficient statistics for the individual unobserved heterogeneity parameters, when these are time invariant. Although sufficient statistics can only be derived in very special cases for the dynamic logit model, the estimation methods proposed by Honoré and Kyriazidou (2000) and Bartolucci and Nigro (2012) are still of CML type and, therefore, follow a fixed-effects approach.

One drawback of the CML method for panel logit models is that it assumes strict exogeneity of the covariates, conditional on unobserved heterogeneity, which is required for consistent estimation of the regression parameters. However, this assumption is likely to be violated because there may be feedback effects from the outcome variable on the future values of the covariates, in which case covariates are said to be predetermined. While in linear models the mainstream approach to overcome this problem is based on instrumental variables (Anderson and Hsiao, 1981; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998) and testing procedures have been developed for heterogeneous panels (Emirmahmutoglu and Kose, 2011; Dumitrescu and Hurlin, 2012), considerably fewer results are available for nonlinear binary panel data models with predetermined covariates. This is particularly true with short binary panel data, where no general solution is yet available despite their relevance in microeconomic applications.

So far, the literature on fixed-effects nonlinear panel data models has focused on bias, score, or likelihood corrections aimed at mitigating the inconsistency of the Maximum Likelihood (ML) estimator that arises from the incidental parameters problem (Neyman and Scott, 1948). This approach, however, gives rise to corrected ML estimators that perform well with many time periods, say at least eight, in finite samples. Notable contributions are those described in Carro (2007), who derived a score correction, Bartolucci et al. (2016), who instead proposed correcting the log-likelihood, and Hahn and Newey (2004), who put forward panel-jackknife and analytical bias corrections later extended by Fernandez-Val (2009) and Hahn and Kuersteiner (2011) to dynamic binary choice models. Among these, Fernandez-Val (2009) explicitly considered predetermined explanatory variables, other than the lagged dependent variable, but the finite-sample properties of the proposed correction are only provided with strictly exogenous covariates.

A different strand of literature has also considered short panel data. Honoré and Lewbel (2002) proposed a semiparametric estimator for the parameters of a binary choice model with predetermined covariates. However, they provided identification conditions when there is a further regressor that is continuous, strictly exogenous, and independent of the individual specific effects. These requirements are often difficult to be fulfilled in practice. Arellano and Carrasco (2003) considered semiparametric random-effects models where covariates are allowed to be predetermined and correlated with the individual specific effects; they proposed a Generalized Method of Moments (GMM) estimator involving the probability distribution of the predetermined covariates (sample cell frequencies for discrete covariates or nonparametric smoothed estimates for continuous covariates) that can, however, be difficult to employ with many explanatory variables. A different approach is taken by Wooldridge (2000), who proposed to specify a joint model for the response variable and the predetermined covariates; the model parameters are estimated by a correlated random-effects approach (Mundlak, 1978; Chamberlain, 1984), so as to account for the dependence between strictly exogenous explanatory variables and individual unobserved effects, combined with a preliminary version of the Wooldridge (2005)'s solution to the initial conditions problem. Although this is a natural strategy, it requires distributional assumptions on the individual unobserved heterogeneity; moreover, it is computationally demanding when the number of predetermined covariates is large.

A strategy similar to that developed by Wooldridge (2000) is adopted by Mosconi and Seri (2006), who adopted ML-based tests for the presence of feedback effects in binary bivariate time-series. They based the estimation and testing strategy on the definition of Granger causality (Granger, 1969), which is typical of the time-series literature, as adapted to the nonlinear panel data setting by Chamberlain (1982) and Florens and Mouchart (1982). While attractive, Mosconi and Seri's approach does not account for individual time-invariant unobserved heterogeneity and is better suited for quite long panels, whereas applications, such as those focused on intertemporal choices in the labor market, poverty traps, and persistence in unemployment, often rely on short time-series and a large number of cross-sectional units resulting from rotated surveys. Furthermore, dealing properly with time-invariant unobserved heterogeneity is crucial for the attainability of the estimation results, since individual-specific effects are often correlated with the covariates of interest. For instance, in dynamic binary choice models, the focus is often on properly detecting the causal effects of past events of the phenomenon of interest, namely the *true* state dependence, as opposed to the persistence generated by permanent individual unobserved heterogeneity (Heckman, 1981).

In this paper, we propose a logit model formulation for a (dynamic) binary fixed T-panel data model that takes into account general forms of feedback effect from the the outcome variable on the future values of the covariates. The logit model parameters can be consistently estimated by the CML method, so as to avoid any parametric assumption on the subject-specific time-invariant unobserved heterogeneity, which is also allowed to be freely correlated with the covariates.

One advantage of our formulation is that it does not require the specification of a joint parametric model for the outcome and predetermined explanatory variables, although as specified in the following, our main result holds exactly under certain regularity conditions on the distribution of such covariates. In fact, the starting point to build the proposed model is the definition of strict exogeneity (Sims, 1972), violations of which correspond to the presence of feedback effects, as stated in terms of conditional independence by Chamberlain (1982) for nonlinear models. The strict exogeneity assumption for nonlinear models requires the specification of the probability distribution of the binary dependent variable at each time occasion (y_t) conditional on past, present, and future values of the covariates (x). If the conditioning set includes the lagged dependent variable (y_{t-1}), then the assumption represents a modification of the Sims' strict exogeneity condition, which is proved to be equivalent to the Granger's noncausality condition for nonlinear models (Chamberlain, 1982).

The proposed model also allows for the inclusion of even a large number of predetermined covariates. Under the logit model, it amounts to augmenting the linear index function with a linear combination of the leads of the predetermined covariates, along with the lags of the binary dependent variable if violations of noncausality are considered. We analytically prove that this augmented linear index function corresponds to the logit for the conditional distribution of y_t given the covariates and future values of x, under the assumption that the distribution of the predetermined covariates belongs to the exponential family with dispersion parameters (Barndorff-Nielsen, 1978). The conditional means of these covariates may depend on individual fixed effects. In the other cases, we assume a linear approximation that proves to be effective, while allowing us to maintain a simple approach. As a consequence, any estimation approach giving rise to a consistent estimator of the parameters of the logit model with strictly exogenous covariates can also be applied to obtain a consistent estimator of the parameters of the proposed logit model.

We study the finite-sample performance of the fixed-effects estimator for the proposed model by means of an extensive simulation study. Specifically, we use the standard CML method for the modified logit model when we investigate violations from strict exogeneity, whereas we rely on the Pseudo-CML (PCML) estimator proposed by Bartolucci and Nigro (2012) for the estimation of the modified dynamic logit model under departures from noncausality. We show that these estimators have good finite-sample properties, even when the required conditions on the distribution of the predetermined covariate are not satisfied, with the exception of the dynamic logit model for very short panels, such as with $T = 4$.

Finally, we consider an empirical application where we investigate the effect of the presence of young children in the family on female labor supply, based on a sample drawn form the Michigan Panel Study of Income Dynamics (PSID). This example has been extensively considered in the literature on feedback effects because of the potential effect that labor force participation exerts on future fertility decisions (see, among others, Chamberlain, 1984; Carrasco, 2001; Arellano and Carrasco, 2003; Mosconi and Seri, 2006). In contrast with some of the results available in the literature, we find no evidence of feedback effects relatively to fertility choices and husband's income. This suggests that relying on a flexible fixed-effects approach, where individual effects are freely allowed to depend on both predetermined and strictly exogenous covariates, without imposing any functional form for this correlation, might help avoiding confusion between a simple misspecification of the unobserved heterogeneity and the presence of feedback effects.

The remainder of the paper is organized as follows. Section 2 introduces some preliminary definitions and notation and Section 3 illustrates the proposed model formulation. Section 4 briefly recalls the CML and PCML estimators for the proposed model, Section 5 outlines the simulation study focusing on its results, and Section 6 reports the estimation results for female labor supply with predetermined fertility decisions. Finally, in Section 7 we provide some major conclusions.

2. Preliminaries

In the following, we first review the definitions of strict exogeneity and non-causality for nonlinear models. Then we illustrate the notation and recall the assumptions underlying the logit models considered for the proposed formulation.

2.1. Definitions

Consider panel data for a sample of n units observed at T occasions through a single explanatory variable x_{it} and binary response y_{it}, with $i = 1, \dots, n$ and $t = 1, \dots, T$, where the response variable is affected by a time-constant unobservable intercept c_i. Also let $\boldsymbol{x}_{i,t_1:t_2} = (x_{it_1}, \dots, x_{it_2})'$ and $\boldsymbol{y}_{i,t_1:t_2} = (y_{it_1}, \dots, y_{it_2})'$ denote the column vectors with elements referred to the period from the t_1-th to the t_2-th occasion, so that $\boldsymbol{x}_i = \boldsymbol{x}_{i,1:T}$ and $\boldsymbol{y}_i = \boldsymbol{y}_{i,1:T}$ are referred to the entire period of observation for sample unit i. Note that here we consider only one covariate to keep the illustration simple, but all definitions and results below naturally extend to the case of more covariates per time occasion.

In this framework, and as illustrated by Chamberlain (1982), assuming that the economic life of every individual begins at time $t = 1$, Sims' definition of strict exogeneity is:

Definition. S - *x is* <u>*strictly exogenous*</u> *with respect to y, given c, if y_{it} is independent of $\boldsymbol{x}_{i,t+1:T}$ conditional on c_i and $\boldsymbol{x}_{i,1:t}$, for all i and t, that is,*

$$p(y_{it}|c_i, \boldsymbol{x}_i) = p(y_{it}|c_i, \boldsymbol{x}_{i,1:t}), \quad i = 1, \dots, n, \ t = 1, \dots, T. \tag{1}$$

Therefore, accommodating violations of S amounts to including leads of the covariates in the regression specification.

If a dynamic model is considered, that is, lags of the dependent variables enter the conditioning set, the above definition becomes a modification of Sims' strict exogeneity assumption, denoted by Chamberlain (1982) as S':

Definition. S' - *x is* <u>*strictly exogenous*</u> *with respect to y, given c and the past responses, if y_{it} is independent of $\boldsymbol{x}_{i,t+1:T}$ conditional on c_i, $\boldsymbol{x}_{i,1:t}$, and $\boldsymbol{y}_{i,1:t-1}$, for all i and t, that is,*

$$p(y_{it}|c_i, \boldsymbol{x}_i, \boldsymbol{y}_{i,1:t-1}) = p(y_{it}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t-1}), \tag{2}$$

for $i = 1, \dots, n$ and $t = 1, \dots, T-1$, where $\boldsymbol{y}_{i,t-1}$ disappears from the conditioning argument for $t = 1$.

Furthermore, Chamberlain (1982) showed that S' is equivalent to Granger non-causality, conditional on the unobserved heterogeneity, which is defined as follows:

Definition. G - *The response (y) does not cause the covariate (x) conditional on the time-fixed effect (c) if $x_{i,t+1}$ is conditionally independent of $\boldsymbol{y}_{i,1:t}$, given c_i and $\boldsymbol{x}_{i,1:t}$, for all i and t, that is,*

$$p(x_{i,t+1}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t}) = p(x_{i,t+1}|c_i, \boldsymbol{x}_{i,1:t}), \tag{3}$$

for $i = 1, \dots, n$ and $t = 1, \dots, T-1$ We provide a proof of equivalence between G and S' in Appendix Appendix A. This proof is related to that provided in Chamberlain (1982).

It is worth noting that accommodating departures from G would require the knowledge and formulation of the model for each time-specific covariate given the the previous covariates and responses. Furthermore, apart from the case $T = 2$, property S' is stronger than S. Then, being equivalent to S', G implies S, but in general S does not imply G. In fact, S is expressed avoiding to condition on the previous responses. Also the proof of this result is provided in Appendix Appendix A.

2.2. Logit models

Consider the general case in which, for $i = 1, \dots, n$ and $t = 1, \dots, T$, we observe a binary response variable y_{it} and a vector of k covariates denoted by $\boldsymbol{x}_{it}$. Then, we extend the previous notation by introducing $\boldsymbol{X}_{i,t_1:t_2} = (\boldsymbol{x}_{it_1}, \dots, \boldsymbol{x}_{it_2})$, with $\boldsymbol{X}_i = \boldsymbol{X}_{i,1:T}$ being the matrix of the covariates for all time occasions. Let us also define the individual matrix $\boldsymbol{W}_{it}$ that, in the following, will be equal to $\boldsymbol{X}_i$ for the static binary choice model, whereas it will also include $\boldsymbol{y}_{i,1:t-1}$ if a dynamic formulation is considered.

The static formulation of a (dynamic) binary choice model assumes that, for all i and t, the binary response y_{it} has conditional distribution

$$p(y_{it}|c_i, \boldsymbol{W}_{it}) = p(y_{it}|c_i, \boldsymbol{w}_{it}), \tag{4}$$

with dependence either on the present values of the explanatory variables, when $\boldsymbol{w}_{it} = \boldsymbol{x}_{it}$, or also on the first lag of the dependent variable, when $\boldsymbol{w}_{it} = (\boldsymbol{x}'_{it}, y_{i,t-1})'$ because a dynamic binary choice model is considered. The latter corresponds to a first-order Markov model for y_{it}. The above conditioning set can be easily enlarged to include further lags of $\boldsymbol{x}_{it}$ (and y_{it}).

Adopting a logit formulation for the conditional probability implies that

$$p(y_{it}|c_i, \boldsymbol{w}_{it}) = \frac{\exp\left[y_{it}\left(c_i + \boldsymbol{w}'_{it}\boldsymbol{\alpha}\right)\right]}{1 + \exp\left(c_i + \boldsymbol{w}'_{it}\boldsymbol{\alpha}\right)}, \tag{5}$$

where the individual-specific intercepts c_i are often considered as nuisance parameters, and $\boldsymbol{\alpha}$ is a vector collecting the parameters of interest. Within the framework of the static logit model, we let $\boldsymbol{\alpha} = \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the vector of regression parameters for the covariates $\boldsymbol{x}_{it}$. If, instead, a dynamic logit model is considered (see Hsiao, 2015, ch. 7, for a review), we let $\boldsymbol{\alpha} = (\boldsymbol{\beta}', \gamma)'$, where γ measures the true state dependence (Heckman, 1981).

It is useful to distinguish the formulation of the conditional distribution of the overall vector of responses for the static logit model from that for the dynamic logit model. In the first case, we have

$$p(\boldsymbol{y}_i|c_i, \boldsymbol{X}_i) = \frac{\exp\left(y_{i+}c_i + \sum_{t=1}^{T} y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta}\right)}{\prod_{t=1}^{T}\left[1 + \exp\left(c_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}\right)\right]}, \tag{6}$$

with $y_{i+} = \sum_{t=2}^{T} y_{it}$ being the *total score*, whereas for the dynamic logit model, the conditional distribution becomes

$$p(\boldsymbol{y}_{i,2:T}|c_i, \boldsymbol{X}_i, y_{i1}) = \frac{\exp\left[y_{i+}c_i + \sum_{t=2}^{T} y_{it}\left(\boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma\right)\right]}{\prod_{t=2}^{T}\left[1 + \exp\left(c_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma\right)\right]}, \tag{7}$$

where the initial observation y_{i1} is considered as given. Fixed-effects formulations have the advantage of not requiring any modeling of the initial observations of the sample (Hsiao, 2015). In fact, the so-called "initial conditions" problem only arises within random-effects models, where an endogeneity issue is posed by the correlation of the lagged dependent variable with the unobserved effects.

Expression (4) embeds assumptions S and S', according to whether a static or dynamic binary choice model is considered, by excluding leads of $\boldsymbol{x}_{it}$ from the probability conditioning set. Therefore, it rules out feedbacks from the response variable to future covariates. The absence of these feedback effects is often a hardly tenable assumption, as when the covariates of interest depend on individual choices. If the covariates are predetermined, as opposed to strictly exogenous, estimation of the model parameters of interest can be severely biased when it is based on eliminating or approximating c_i with quantities depending on the entire observed history of covariates (Mundlak, 1978; Chamberlain, 1984; Wooldridge, 2005).

3. Proposed model formulation

As stated in Section 2, dealing with violations of S and S' amounts to proposing a generalization of the static or dynamic binary choice model based on assumption (4). In order to derive the proposed model, which is a binary choice with feedback effects, we specify the probability of y_{it} conditional on the individual intercept and on $\boldsymbol{W}_{it}$ as

$$p(y_{it}|c_i, \boldsymbol{W}_{it}) = p(y_{it}|c_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1}), \tag{8}$$

retaining the assumption that previous covariates and responses before $y_{i,t-1}$ do not affect y_{it}. Note that, differently from (4), the conditioning set at the rhs of (8) includes the first-order leads of $\boldsymbol{x}_{it}$, therefore accommodating violations of S and S'. The formulation can easily be extended to include an arbitrary number H of leads collected in $\boldsymbol{X}_{i,t+1:t+H}$, with $H \leq T-2$ for the static and $H \leq T-3$ for the dynamic binary choice model, so that we maintain at least two observations, which is necessary for identification (see Section 4). However, we do not explicitly consider the extension to an arbitrary number of leads because, while being rather obvious, it strongly complicates the exposition. In this regard note that Chamberlain (1984) reported an empirical example where the linear index function of a logit model corresponds to the lhs of S in (1), where all the available lags and leads of $\boldsymbol{x}_{it}$ are used. However, this specification is valid only when $t = 1$ is the beginning of the subject's economic life. We do not make the same assumption here.

At this point, it is worth to stress that if we are in presence of violations of S and S', any estimation approach that requires strict exogeneity of the covariates will produce an inconsistent estimator of the parameters of the following logit model:

$$p(y_{it}|c_i, \boldsymbol{w}_{it}) = \frac{\exp\left[y_{it}(c_i + \boldsymbol{w}_{it}'\boldsymbol{\vartheta})\right]}{1 + \exp(c_i + \boldsymbol{w}_{it}'\boldsymbol{\vartheta})}. \tag{9}$$

This model neglects feedback effects even though covariates are predetermined, as per expression (8). Here $\boldsymbol{\vartheta}$ collects the parameters of interest: within the static framework, $\boldsymbol{\vartheta} = \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is the vector of regression parameters, otherwise $\boldsymbol{\vartheta} = (\boldsymbol{\mu}', \delta)'$, where δ represents the true state dependence. If instead S or S' hold as in (4), then $\boldsymbol{\vartheta} = \boldsymbol{\alpha}$ and (5) is the same as (9). As already mentioned in Section 1, Wooldridge (2000) proposed to set up a multivariate model for (9) and the predetermined covariates. On the contrary, the formulation proposed below has the advantage of not requiring specification of a joint parametric model for the outcome and predetermined explanatory variables. It is also worth recalling that estimation approaches not requiring strict exogeneity for consistency typically suffer from other sources of bias. We are referring to the ML estimator when it is used to estimate a random-effects model in presence of violations of the distributional assumptions or to a fixed-effects model with unobserved heterogeneity due to the incidental parameters problem.

The proposed formulation of a specific model for (8) and consequent identification of its parameters require further assumptions that lead to the formulation here proposed. In particular, we rely on the logit formulation

$$p(y_{it}|d_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1}) = \frac{\exp\left[y_{it}\left(d_i + \boldsymbol{w}_{it}'\boldsymbol{\vartheta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu}\right)\right]}{1 + \exp\left(d_i + \boldsymbol{w}_{it}'\boldsymbol{\vartheta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu}\right)}. \tag{10}$$

Here the individual effect is denoted by d_i, as a result of the identifying assumptions that will be made below. Following the suggestion in Wooldridge (2010, Sec. 15.8.2), a test for strict exogeneity and/or noncausality can be derived by specifying a model of this type. In fact, the null hypothesis $H_0 : \boldsymbol{\nu} = \boldsymbol{0}$, where $\boldsymbol{0}$ is a column vector of zeros of suitable dimension, corresponds to condition S or S', according to whether a static or dynamic formulation is considered. Therefore, under H_0, the proposed model corresponds to the static or dynamic logit model, with $\boldsymbol{\vartheta} = \boldsymbol{\alpha}$ and $d_i = c_i$.

We show that under a particular but very relevant case, the formulation in (10) represents a logit model with feedback effects. The justification is based on the following arguments. First of all, denote the conditional density of the distribution of the covariate vector $\boldsymbol{x}_{i,t+1}$ as

$$f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{X}_{i,1:t}, \boldsymbol{y}_{i,1:t}) = f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it}), \quad t = 1, \dots, T-1, \tag{11}$$

where $\boldsymbol{\xi}_i$ is a column vector of time-fixed effects and the presence of y_{it} in the conditioning set allows for feedback effects. Equation (11) also depicts the conditional independence of $\boldsymbol{x}_{i,t+1}$ from $\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{i,t-1}$ and $y_{i1}, \dots, y_{i,t-1}$ given $\boldsymbol{x}_{it}, y_{it}$, which can however be relaxed by including more lags of $\boldsymbol{x}_{it}$ and y_{it}.

Then the logit for the distribution y_{it} conditional on c_i, $\boldsymbol{\xi}_i$, $\boldsymbol{w}_{it}$, and $\boldsymbol{x}_{i,t+1}$ is

$$\log\frac{p(y_{it}=1|c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})}{p(y_{it}=0|c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})} = \log\frac{f(y_{it}=1, \boldsymbol{x}_{i,t+1}|c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it})}{f(y_{it}=0, \boldsymbol{x}_{i,t+1}|c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it})}$$

$$= \log \frac{p(y_{it} = 1 | c_i, \boldsymbol{w}_{it}) f(\boldsymbol{x}_{i,t+1} | \boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it} = 1)}{p(y_{it} = 0 | c_i, \boldsymbol{w}_{it}) f(\boldsymbol{x}_{i,t+1} | \boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it} = 0)}, \tag{12}$$

where the presence of time-fixed effects in the conditioning sets for y_{it} and $\boldsymbol{x}_{it}$ derives from equations (8) and (11). Furthermore, we assume that the probability of y_{it} conditional on c_i and $\boldsymbol{w}_{it}$ has the logit formulation in (9), so that the above expression becomes

$$\log \frac{p(y_{it} = 1 | c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})}{p(y_{it} = 0 | c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})} = c_i + \boldsymbol{w}'_{it}\boldsymbol{\vartheta} + \log \frac{f(\boldsymbol{x}_{i,t+1} | \boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it} = 1)}{f(\boldsymbol{x}_{i,t+1} | \boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it} = 0)}.$$

The main point now is how to deal with the components involving the ratio between the conditional density of $\boldsymbol{x}_{i,t+1}$ for $y_{it} = 0$ and $y_{it} = 1$. Suppose that the conditional distribution of $\boldsymbol{x}_{i,t+1}$ belongs to the exponential family formulated as

$$f(\boldsymbol{x}_{i,t+1} | \boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it} = z) = \frac{\exp[\boldsymbol{x}'_{i,t+1}(\boldsymbol{\xi}_i + \boldsymbol{\eta}_z)] h(\boldsymbol{x}_{i,t+1}; \boldsymbol{\sigma})}{K(\boldsymbol{\xi}_i + \boldsymbol{\eta}_z; \boldsymbol{\sigma})}, \tag{13}$$

with $t = 1, \ldots, T-1$ and $z = 0, 1$, where $h(\boldsymbol{x}_{i,t+1})$ is an arbitrary strictly positive function, possibly depending on suitable dispersion parameters $\boldsymbol{\sigma}$, and $K(\cdot)$ is the normalizing constant. Note that this structure also covers the case of $\boldsymbol{x}_{i,t+1}$ depending on time-fixed effects through $\boldsymbol{\xi}_i$. The following result may be simply proved.

Theorem 1. *Assumptions* (9) *and* (13) *imply that*

$$\log \frac{p(y_{it} = 1 | c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})}{p(y_{it} = 0 | c_i, \boldsymbol{\xi}_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})} = \log \frac{p(y_{it} = 1 | d_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})}{p(y_{it} = 0 | d_i, \boldsymbol{w}_{it}, \boldsymbol{x}_{i,t+1})}$$

$$= d_i + \boldsymbol{w}'_{it}\boldsymbol{\vartheta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu},$$

where $d_i = c_i + \log K(\boldsymbol{\xi}_i + \boldsymbol{\eta}_1; \boldsymbol{\sigma}) - \log K(\boldsymbol{\xi}_i + \boldsymbol{\eta}_0; \boldsymbol{\sigma})$ *and* $\boldsymbol{\nu} = \boldsymbol{\eta}_1 - \boldsymbol{\eta}_0$, *and then model* (10) *holds.*

Corollary 1. *Under the conditions of Theorem 1, any estimation approach giving rise to a consistent estimator of the parameters of the logit model* (5), *when* S *or* S' *hold, can also be applied to obtain a consistent estimator of the parameters of the proposed logit model* (10) *and then of* (9) *in case of violations of* S *or* S'.

Corollary 1 simply states that, if Theorem 1 is verified, the parameters of model (9) can be consistently estimated even if S or S' do not hold by implementing a strategy based on an estimator that requires strict exogeneity of covariates applied to model (10). Consider for instance the CML approach, which produces a consistent estimator of $\boldsymbol{\beta}$ in (5) if S holds. On the contrary, the CML estimator will not be consistent for $\boldsymbol{\vartheta} = \boldsymbol{\mu}$ in (9) because of the violations of S, if we exclude the leads of the covariates. A consistent CML estimator for $\boldsymbol{\mu}$ can instead be obtained if the model specification includes the future values of the predetermined covariates as in (10), provided that (13) holds.

Two cases satisfying (13) are for continuous covariates having multivariate normal distribution with common variance-covariance matrix and for binary covariates with distribution based on a logit parametrization. More precisely, in the first case we suppose that

$$\boldsymbol{x}_{i,t+1}|c_i, \boldsymbol{x}_{it}, y_{it} = z \sim N(\boldsymbol{\zeta}_i + \boldsymbol{\mu}_z, \boldsymbol{\Sigma});$$

then (13) holds with $\boldsymbol{\xi}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\zeta}_i$ and $\boldsymbol{\eta}_z = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_z$, $z = 0, 1$, where the upper (lower) triangular part of $\boldsymbol{\Sigma}$ goes in $\boldsymbol{\psi}$. In the second case, we suppose that given $\boldsymbol{\xi}_i$, $\boldsymbol{X}_{it}$, and $y_{it} = z$, the elements of $\boldsymbol{x}_{i,t+1}$ are conditionally independent, with the j-th element having a Bernoulli distribution with success probability

$$\frac{\exp(\xi_{ij} + \eta_{zj})}{1 + \exp(\xi_{ij} + \eta_{zj})}, \quad j = 1, \ldots, k,$$

where k is the number of covariates.

There may be several situations where (13) does not hold. In these cases, we anyway assume a linear approximation for the ratio between the conditional density of $\boldsymbol{x}_{i,t+1}$ for $y_{it} = 0$ and $y_{it} = 1$ in (12), which is the most natural solution to maintain an acceptable level of simplicity. Two examples are investigated by simulation in Section 5, in which the predetermined covariate is allowed to depend on a time-varying explanatory variable and is generated according to a probit model. The results of these simulation exercises suggest that the specification of the log-odds ratio described in Theorem 1 provides a good approximation in presence of these violations.

Another common situation of violation (13) in empirical applications is due to the presence of some persistence characterizing the predetermined covariate, which deserves a special mention. In fact, there are several components that can give rise to a time dependence in $\boldsymbol{x}_{i,t+1}$, some of which can be handled within the set of hypotheses considered. Specifically, the presence of time-invariant unobserved heterogeneity and feedback effects is explicitly taken into account in (13), whereas time-varying explanatory variables can be included in the model for $\boldsymbol{x}_{i,t+1}$ considering expression (12) as a linear approximation.

The presence of *true* state dependence, instead, cannot be easily accounted for by the proposed approach, as it causes an identification problem in (12). This case corresponds to $\boldsymbol{x}_{i,t+1}$ following an autoregressive process. For instance if $\boldsymbol{x}_{i,t+1}$ follows an AR(1) process, then $K(\cdot)$ in Theorem 1 will include $\boldsymbol{x}_{it}$, which will then be part of the unobserved effect d_i. This gives rise to an endogeneity problem due to relevant omitted variables, which makes the CML and PCML estimators for the proposed model inconsistent and with a nonnegligible finite-sample bias, thereby preventing us from considering the proposed linear approximation as effective. As a matter of fact, the proposed approach in this case still represents a viable tool only for testing S and S' (simulation results are available upon request from the Authors). This problem, however, is substantially downsized if we consider the conceptually similar case of error terms following an autoregressive process, which is likely to occur in practice. We illustrate the finite-sample results for AR(1) errors in the model for $\boldsymbol{x}_{i,t+1}$ in Section 5, which are in line with those of the scenarios where Theorem 1 holds.

For the following developments, it is convenient to derive the conditional distribution of the entire vector of responses, which holds under the extended logit formulation (10). It is also useful to separate the static from the dynamic logit model, so as to clarify the differences in the time occasions used and treatment of initial conditions.

The conditional distribution of the overall vector of responses under the static logit model directly compares with (6). For all i, the distribution at issue is

$$p(\boldsymbol{y}_{i,1:T-1}|d_i, \boldsymbol{X}_i, y_{iT}) = \frac{\exp\left[y_{i+}d_i + \sum_{t=1}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right)\right]}{\prod_{t=1}^{T-1}\left[1 + \exp\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right)\right]}, \qquad (14)$$

where $y_{i+} = \sum_{t=1}^{T-1} y_{it}$. In particular, model (14) reduces to the static logit (6) under the null hypothesis of strict exogeneity, namely $H_0 : \boldsymbol{\nu} = \boldsymbol{0}$, if the corresponding probability is conditional on y_{iT} and with different individual intercepts. Moreover, the overall vector of responses for the dynamic logit model is related to (7) and has distribution

$$p(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = \frac{\exp\left[y^*_{i+}d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\delta\right)\right]}{\prod_{t=2}^{T-1}\left[1 + \exp\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\delta\right)\right]}, \qquad (15)$$

where $y^*_{i+} = \sum_{t=2}^{T-1} y_{it}$. Model (15) reduces to the dynamic logit (7) under $H_0 : \boldsymbol{\nu} = \boldsymbol{0}$ and the same conditions expressed above.

4. Estimation

In this section, we illustrate the CML and the PCML estimators for the proposed models.

4.1. CML estimator

Conditional inference for the static logit model is based on the conditional likelihood given the total scores, which are suitable sufficient statistics for the incidental parameters (Chamberlain, 1980). In general, the parameters of the proposed models can be estimated pursuing either a fixed-effects or a (correlated) random-effects strategy (Mundlak, 1978; Chamberlain, 1984; Wooldridge, 2005). The latter, however, only allows the unobserved heterogeneity to be correlated with strictly exogenous covariates, while requiring the predetermined covariates in $\boldsymbol{x}_{it}$ to be independent of d_i. As this assumption may often be hardly tenable, we focus on fixed-effects estimation approaches. The probability in (14), conditional on y_{i+}, becomes

$$p(\boldsymbol{y}_{i,1:T-1}|y_{i+},\boldsymbol{X}_i,y_{iT}) = \frac{\exp\left[\sum\limits_{t=1}^{T-1} y_{it}\left(\boldsymbol{x}_{it}'\boldsymbol{\mu}+\boldsymbol{x}_{i,t+1}'\boldsymbol{\nu}\right)\right]}{\sum\limits_{\boldsymbol{z}_{1:T-1}}\exp\left[\sum\limits_{t=1}^{T-1} z_t\left(\boldsymbol{x}_{it}'\boldsymbol{\mu}+\boldsymbol{x}_{i,t+1}'\boldsymbol{\nu}\right)\right]}, \qquad (16)$$

which no longer depends on d_i and where the sum at the denominator is extended to all possible response configurations $\boldsymbol{z}_{1:T-1} = (z_1,\ldots,z_{T-1})'$ such that $z_+ = y_{i+}$, where $z_+ = \sum\limits_{t=1}^{T-1} z_t$. The parameter vector $\boldsymbol{\phi} = (\boldsymbol{\mu}',\boldsymbol{\nu}')'$ is estimated by maximizing, through a Newton-Raphson algorithm, the conditional log-likelihood corresponding to (16), which can be written as

$$\begin{aligned}\ell(\boldsymbol{\phi}) &= \sum_i 1\{0<y_{it}<T-1\}\ell_i(\boldsymbol{\phi}),\\ \ell_i(\boldsymbol{\phi}) &= \log p(\boldsymbol{y}_{i,1:T-1}|y_{i+},\boldsymbol{X}_i,y_{iT}).\end{aligned}$$

The resulting vector $\hat{\boldsymbol{\phi}} = (\hat{\boldsymbol{\mu}}',\hat{\boldsymbol{\nu}}')'$ is the CML estimate. Expressions for the score vector and information matrix can be derived using the standard theory on the regular exponential family (Barndorff-Nielsen, 1978) and are implemented in package `cquad` (Bartolucci and Pigini, 2017), available in `R` and `Stata`, that we suggest for the application of the estimation method.

Under mild regularity conditions, concerning essentially the structure of the covariate matrix so as to avoid problems of singularity, the CML estimator is consistent as n grows to infinity with fixed T. Moreover, it has an asymptotic Normal distribution with variance-covariance matrix that may estimated in the usual way on the basis of the Hessian of $\ell(\boldsymbol{\phi})$. From this matrix it is also possible to obtain standard errors for the parameter estimates. An illustration of these properties may be found in textbooks such as Hsiao (2015).

4.2. PCML estimator

If a dynamic logit model is considered, sufficient statistics for the individual intercepts can only be derived in absence of covariates with $T = 3$ (Chamberlain, 1985). In presence of covariates, a weighted conditional log-likelihood may be used for inference, although the estimator is consistent only under certain regularity conditions on the distribution of the covariates and the rate of convergence is slower than $\sqrt{n}$ (Honoré and Kyriazidou, 2000). These shortcomings have been overcome by Bartolucci and Nigro (2012), who proposed to approximate the dynamic logit by a Quadratic Exponential (QE) model (Cox, 1972; Bartolucci and Nigro, 2010), which admits sufficient statistics for the incidental parameters and has the same interpretation as the dynamic logit model in terms of log-odds ratio. Bartolucci and Nigro (2012) also proposed to adopt a PCML estimator for the model parameters. This estimator is consistent in absence of true state dependence and has a negligible bias in case of even strong state dependence.

The approximating model used in Bartolucci and Nigro (2012), and here adapted for (15), is derived by taking a linearization of the log-probability of the latter, that is,

$$\begin{aligned}\log p(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) \quad = \quad & y_{i+}^* d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}_{it}'\boldsymbol{\mu} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\delta\right) \\ & - \sum_{t=2}^{T-1} \log\left[1 + \exp\left(d_i + \boldsymbol{x}_{it}'\boldsymbol{\mu} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\delta\right)\right]. \quad (17)\end{aligned}$$

The term that is nonlinear in the parameters is approximated by a first-order

Taylor series expansion around $d_i = \bar{d}_i$, $\boldsymbol{\mu} = \bar{\boldsymbol{\mu}}$, $\boldsymbol{\nu} = \bar{\boldsymbol{\nu}}$, and $\delta = 0$, leading to

$$\begin{aligned}
&\sum_{t=2}^{T-1} \log\left[1 + \exp\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\delta\right)\right] \\
&\qquad \approx \sum_{t=2}^{T-1} \log\left[1 + \exp\left(\bar{d}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\mu}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)\right] \\
&\quad + \sum_{t=2}^{T-1} q_{it}\left[d_i - \bar{d}_i + \boldsymbol{x}'_{it}\left(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}\right) + \boldsymbol{x}'_{i,t+1}\left(\boldsymbol{\nu} - \bar{\boldsymbol{\nu}}\right)\right] + \sum_{t=2}^{T-1} q_{it} y_{i,t-1}\delta, \qquad (18)
\end{aligned}$$

where

$$q_{it} = \frac{\exp\left(\bar{d}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\mu}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)}{1 + \exp\left(\bar{d}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\mu}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)}.$$

Since only the last sum in (18) depends on $\boldsymbol{y}_{i,2:T-1}$, we can substitute this sum in (17) and obtain the approximation of the joint probability (15) that gives the QE model

$$\begin{aligned}
&p^{\dagger}(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) \\
&\quad = \frac{\exp\left[y^*_{i+}d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(y_{it} - q_{it})y_{i,t-1}\delta\right]}{\sum_{\boldsymbol{z}_{2:T-1}} \exp\left[z^*_{+}d_i + \sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(z_t - q_{it})z_t\delta\right]}. \qquad (19)
\end{aligned}$$

The sum at the denominator of the previous expression ranges over all possible binary vectors $\boldsymbol{z}_{2:T-1} = (z_2, \ldots, z_{T-1})'$, with $z^*_+ = \sum_{t=2}^{T-1} z_t$ and $z_1 = y_{i1}$.

The joint probability in (19) is closely related to the probability of the response configuration $\boldsymbol{y}_{i,2:T-1}$ in the true model. In particular: (*i*) expressions (19) and (15) correspond to the same logit model in absence of state dependence; (*ii*) in both models, y_{it} is conditionally independent of $\boldsymbol{y}_{i,1:t-2}$ given d_i, $\boldsymbol{X}_i$, and $y_{i,t-1}$; (*iii*) the parameter δ has the same interpretation in terms of log-odds ratio between the responses y_{it} and $y_{i,t-1}$. These results can be proved along the lines of Theorem 1 in Bartolucci and Nigro (2010).

The nice feature of the QE model in (19) is that it admits sufficient statistics for the incidental parameters d_i, which are the total scores y^*_{i+} for $i = 1, \ldots, n$.

Under the approximating model, the probability of $\boldsymbol{y}_{i,2:T-1}$, conditional on $\boldsymbol{X}_i$, y_{i1}, y_{iT}, and y^*_{i+}, is then

$$
\begin{aligned}
p^\dagger\left(\boldsymbol{y}_{i,2:T-1}|y^*_{i+}, \boldsymbol{X}_i, y_{i1}, y_{iT},\right) &= \frac{p^\dagger(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT})}{p^\dagger(y^*_{i+}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT})} \\
&= \frac{\exp\left[\sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(y_{it} - q_{it})y_{i,t-1}\delta\right]}{\sum_{\substack{\boldsymbol{z}_{2:T-1}\\ z^*_+ = y^*_{i+}}} \exp\left[\sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(z_t - q_{it})z_{t-1}\delta\right]}, \qquad (20)
\end{aligned}
$$

which no longer depends on d_i and where the sum at the denominator is extended to all possible response configurations $\boldsymbol{z}_{2:T-1}$ such that $z^*_+ = y^*_{i+}$. he denominator is

$$
p^\dagger(y^*_{i+}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = \frac{\sum_{\substack{\boldsymbol{z}_{2:T-1}\\ z^*_+ = y^*_{i+}}} \exp\left[z^*_+ d_i + \sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(z_t - q_{it})z_t\delta\right]}{\sum_{\boldsymbol{z}_{2:T-1}} \exp\left[z^*_+ d_i + \sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\boldsymbol{\mu} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(z_t - q_{it})z_t\delta\right]},
$$

which clarifies that $y^*_{i+}d_i$ in (19) and $z^*_+ d_i$ cancel out.

The formulation of the conditional log-likelihood based on (20) relies on the fixed quantities q_{it}, which are based on a preliminary estimation of the parameters associated with the covariate and the individual effects. Let $\boldsymbol{\phi} = (\boldsymbol{\mu}', \boldsymbol{\nu}')'$ and $\boldsymbol{\theta} = (\boldsymbol{\phi}', \boldsymbol{\delta}')'$. The estimation approach is based on two-steps:

1. Preliminary estimates of the parameters needed to compute q_{it} are obtained by maximizing the conditional log-likelihood

$$
\begin{aligned}
\ell(\bar{\boldsymbol{\phi}}) &= \sum_{i=1}^{n} 1\{0 < y_{it} < T-2\}\ell_i(\bar{\boldsymbol{\phi}}), \\
\ell_i(\bar{\boldsymbol{\phi}}) &= \log \frac{\exp\left[\sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\bar{\boldsymbol{\mu}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)\right]}{\sum_{\substack{\boldsymbol{z}_{2:T-1}\\ z^*_+ = y^*_{i+}}} \exp\left[\sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\bar{\boldsymbol{\mu}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)\right]},
\end{aligned}
$$

which can be performed by a Newton-Raphson algorithm.

2. The parameter vector $\boldsymbol{\theta}$ is estimated by maximizing the conditional log-likelihood corresponding to (20), which can be written as

$$\begin{aligned}\ell^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \sum_i 1\{0 < y_{it} < T-2\}\ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}),\\ \ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \log p^{\dagger}_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}(\boldsymbol{y}_{i,2:T-1}|y^*_{i+}, \boldsymbol{X}_i, y_{i1}, y_{iT}).\end{aligned}$$

The resulting $\hat{\boldsymbol{\theta}}$ is the PCML estimate.

Similarly to the conditional log-likelihood on which the CML estimator is based, function $\ell^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ may be maximized by a Newton-Raphson algorithm using the score vector and observed information matrix, which are computed on the basis of the standard theory on the regular exponential family. Moreover, for the application of the proposed PCML we again suggest to adopt package `cquad`. Regarding the asymptotic properties of this estimator we recall that, essentially, it has the same asymptotic properties of the CML estimator in absence of state dependence, whereas with state dependence it converges in probability, as n grows to infinity with T fixed, to a pseudo-true parameter that is reasonably close to the true parameter value. The asymptotic normal distribution also holds when there is state dependence, although the estimation of the variance-covariance matrix is more complex than for the CML estimator, as it has to account for the estimated quantities in step 1.

Following Bartolucci and Nigro (2012), the expression for the variance-covariance matrix estimator is based on a GMM approach (Hansen, 1982). In fact, the PCML estimator can be seen as the solution of the system of equations

$$\boldsymbol{g}(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) = \sum_{i=1}^{n} 1\{0 < y^*_{i+} < T-2\}\boldsymbol{g}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) = \boldsymbol{0},$$

where

$$\boldsymbol{g}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\nabla}_{\bar{\boldsymbol{\phi}}}\ell_i(\bar{\boldsymbol{\phi}}) \\ \boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) \end{pmatrix},$$

with $\ell_i(\bar{\boldsymbol{\phi}})$ and $\ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ defined in previous steps 1 and 2, respectively. Denoting the solution to this equation as $(\tilde{\boldsymbol{\phi}}', \hat{\boldsymbol{\theta}}')'$, the asymptotic variance-covariance matrix can be estimated by

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

where

$$\boldsymbol{S}(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) = \sum_{i=1}^{n} 1\{0 < y_{i+}^* < T-2\} \boldsymbol{g}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) \boldsymbol{g}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta})',$$

and

$$\boldsymbol{H}(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) = \sum_{i=1}^{n} 1\{0 < y_{i+}^* < T-2\} \boldsymbol{H}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}).$$

In the previous expression, we have that

$$\boldsymbol{H}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) = \begin{pmatrix} \nabla_{\bar{\boldsymbol{\phi}}\bar{\boldsymbol{\phi}}} \ell_i(\bar{\boldsymbol{\phi}}) & \boldsymbol{O} \\ \nabla_{\boldsymbol{\theta}\bar{\boldsymbol{\phi}}} \ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) & \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) \end{pmatrix}$$

is the derivative of $\boldsymbol{g}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta})$ with respect to $(\bar{\boldsymbol{\phi}}', \boldsymbol{\theta}')'$, with $\boldsymbol{O}$ denoting a matrix of zeros of suitable dimension. For the computation of the variance-covariance matrix estimator, analytical expressions are used for $\nabla_{\bar{\boldsymbol{\phi}}\bar{\boldsymbol{\phi}}} \ell_i(\bar{\boldsymbol{\phi}})$ and $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$, which can easily be derived on the basis of the formulations given in steps 1 and 2, respectively, whereas we rely on numerical differentiation for the evaluation of $\nabla_{\boldsymbol{\theta}\bar{\boldsymbol{\phi}}} \ell_i^{\dagger}(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$.

5. Simulation study

In this section, we describe the design and illustrate the results of the simulation study of the finite-sample properties of the CML and PCML estimators for the proposed model formulations. The study also compares the PCML estimator with the random-effects ML estimator for the joint model for the binary dependent variable and predetermined covariate, as proposed by Wooldridge (2000).

5.1. Simulation design

The simulation study is based on samples drawn from a logit model with one explanatory variable x_{it} possibly predetermined, one strictly exogenous variable z_{it}, and individual unobserved heterogeneity. The model for the response variable assumes that

$$y_{it} = 1\{c_i + x_{it}\mu - 0.5z_{it} + \varepsilon_{it} \geq 0\}, \quad i = 1, \dots, n,\ t = 1, \dots, T, \tag{21}$$

where the error terms ε_{it} follow a logistic distribution with zero mean and variance equal to $\pi^2/3$, while the individual specific intercepts c_i are allowed to be correlated with x_{it} and z_{it}.

The strictly exogenous covariate z_{it} is generated as

$$z_{it} = \xi_i + v_{it}, \quad i = 1, \dots, n,\ t = 1, \dots, T,$$

where $v_{it} \sim N(0, \pi^2/3)$. Moreover, for $i = 1, \dots, n$, the explanatory variable x_{i1} is obtained from a third-degree polynomial function of $\xi_i + u_{i1}$, with $u_{i1} \sim N(0, \pi^2/3)$, whereas for $t = 2, \dots, T$, it is assumed that

$$\begin{aligned} x_{it} &= \xi_i + z_{it}\psi + y_{i,t-1}\eta + u_{it}, \\ u_{it} &= u_{i,t-1}\rho + \omega_{it}, \end{aligned}$$

with $\omega_{it} \sim N(0, \pi^2/3)$ and where parameter η governs the violation of S, stated in Section 2, and it takes value $\eta = 0$ under the assumption of strict exogeneity, with $\eta \neq 0$ otherwise.

For $i = 1, \dots, n$, the individual intercepts c_i and ξ_i are derived as

$$\begin{aligned} c_i &= \frac{1}{T}\sum_{t=1}^{4} u_{it}, \\ \xi_i &= 0.5\, c_i + \sqrt{0.75}\, \tau_{it}, \end{aligned} \tag{22}$$

where $\tau_{it} \sim N(0, 1)$. In this way, the generating model admits correlation between the covariates and the individual-specific intercepts; it also admits dependence between the unobserved heterogeneity in both processes for y and x. Notice that the simulation design implicitly assumes that the only source of contemporaneous endogeneity, namely the reverse causality between x_{it} and y_{it}, is completely captured by the correlation between the individual specific intercepts in the two processes.

In this framework based on generating model (21), different scenarios are considered; under each of these scenarios, violations of noncausality are examined by setting $\eta = -1$, compared with the same scenarios with $\eta = 0$. The sample sizes considered are $n = 100, 250, 500, 1000$ for $T = 4, 8, 12$ time occasions, with results based on a number of Monte Carlo replications equal to $1,000$. In the first scenario, which we refer to as Experiment 1, we let $\mu = 0$, $\psi = 0$, and $\rho = 0$ so that, also due to the linear model specification for x_{it}, assumption (13) is satisfied and the assumptions of Theorem 1 hold. The simulation study includes four additional experiments, whose designs are as follows:

- Experiment 2: as Experiment 1 with $\mu = -1$;
- Experiment 3: as Experiment 2 with $\psi = 0.5$;
- Experiment 4: as Experiment 2 with $\rho = 0.25$.

Additional scenarios are the following:

- Experiment 5: as Experiment 2 with a binary predetermined covariate, so that for $i = 1, \dots, n$, we have

$$x_{it} = 1\{\xi_i + \eta y_{i,t-1} + u_{it} \geq 0\}, \quad t = 2, \dots, T,$$

  and $p(x_{it} = 1|\xi_i, y_{i,t-1}) = \Phi(\xi_i + \eta y_{i,t-1})$;
- Experiment 6: as Experiment 2 with the inclusion of the lagged dependent variable in (21), so that samples are drawn from the dynamic logit model

based on assuming that, for $i = 1, \dots, n$,

$$y_{it} = 1\{c_i + \mu x_{it} - 0.5 z_{it} + \delta y_{i,t-1} + \varepsilon_{it} \geq 0\}, \quad t = 2, \dots, T,$$

where $\delta = 1$ and with initial condition

$$y_{i1} = 1\{c_i - 0.5 x_{i1} - z_{i1} + \varepsilon_{i1} \geq 0\}.$$

In this setting, violations from noncausality S' are considered through the following scenarios:

- Experiment 7: as Experiment 6 with $\psi = 0.5$;
- Experiment 8: as Experiment 6 with $\rho = 0.25$;
- Experiment 9: as Experiment 6 with a binary predetermined covariate, as in Experiment 5.

Experiment 6 is motivated by economic applications in which the parameter of interest is δ, measuring the state dependence, further to the regression coefficient μ. Moreover, the chosen values for μ, δ, and η are consistent with likely situations in practice that are related to the feedback effect of employment on future child birth when analyzing female labor supply (see Section 6 and also Mosconi and Seri, 2006, for a related application). In Experiments 6 to 9 the simulation study is limited to sample sizes 500 and 1000, due to convergence problem that may arise because of the small number of subjects contributing to the log-likelihood with a high degree of state dependence. In Experiments 3-5 and 7-9, assumption (13) does not hold and the model formulated in Theorem 1 is an approximation. More specifically, in Experiments 3-4 and 7-8 the covariate x_{it} is allowed to depend on a time-varying explanatory variable or to be a function of an AR(1) error term, whereas in Experiments 5 and 9 the distribution of x_{it} does not belong to the exponential family.

5.2. Simulation results

In this section, we describe the results of our simulation study based on Experiments 1 to 9. Under each of the first five of these scenarios, we investigated the finite-sample performance of the CML estimator for the proposed formulation (14) in two cases representing strict exogeneity corresponding to property S described in Section 2 and its violation: CML_1 denotes the CML estimator for the parameters in (14); CML_0 denotes the estimator of (6) under the constraint $\nu = 0$ and with probability conditioned on y_{iT}. Under Experiments 6 to 9, we estimated the model parameters by PCML, with PCML_0 and PCML_1 denoting the hypotheses of noncausality corresponding to S' and its violation.

For each estimator, we report the mean bias, the median bias, the root-mean square error (RMSE), the median absolute error (MAE), as in Honoré and Kyriazidou (2000), and the t-tests at the 5% nominal size for the null hypothesis of μ, and δ in Experiment 5, being equal to the value set in each scenario. Finally we report the t-tests at the 5% nominal size for noncausality, $H_0 : \nu = 0$. We expect CML_0 (PCML_0) to yield biased estimators when $\eta \neq 0$ because, according to S (S'), the lead of x_{it} is omitted from the model specification. We limit the discussion to the estimation of μ, and possibly δ, which are likely to be the parameters of main interest in applications. Results concerning the other model parameters are available upon request.

Tables 1 and 2 summarize the simulation results for our benchmark design with $\mu = 0$ and $\mu = -1$, respectively. With $\eta = 0$, that is, in absence of feedback effects, the mean bias and median bias are always negligible, except when $n = 100$ and $T = 4$, whereas the MAE and RMSE decrease with both n and T for the two models considered. The same considerations hold for CML_1 when $\eta = -1$, whereas the CML estimators of μ denoted by CML_0 is biased and leads to misleading inference, although this pattern is alleviated for $T = 8, 12$. The t-test for $H_0 : \nu = 0$ always attains its nominal size and exhibits strong empirical power in all scenarios with $\eta = -1$, provided T is greater than 4 if $n = 100$.

Tables 3, 4, and 5 report the simulation results for two departures from the

Table 1: Simulation results from Experiment 1: CML estimator, $\mu = 0$, $\psi = 0$, $\rho = 0$, normally distributed covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 100, \quad T = 4$						
CML_1	0.001	0.096	0.004	0.063	0.044	0.055	-0.003	0.098	-0.001	0.066	0.056	0.848
CML_0	0.000	0.089	0.002	0.059	0.040		0.079	0.118	0.078	0.084	0.159	
						$n = 100, \quad T = 8$						
CML_1	-0.003	0.053	-0.003	0.036	0.061	0.058	-0.004	0.053	-0.005	0.035	0.064	1.000
CML_0	-0.002	0.052	-0.004	0.035	0.057		0.037	0.063	0.036	0.043	0.136	
						$n = 100, \quad T = 12$						
CML_1	-0.002	0.038	-0.002	0.027	0.038	0.067	-0.003	0.040	-0.002	0.027	0.053	1.000
CML_0	-0.002	0.038	-0.002	0.026	0.039		0.024	0.045	0.025	0.031	0.102	
						$n = 250, \quad T = 4$						
CML_1	0.000	0.060	0.002	0.039	0.051	0.045	0.003	0.057	0.001	0.040	0.048	0.999
CML_0	0.000	0.056	0.000	0.037	0.056		0.081	0.097	0.083	0.083	0.342	
						$n = 250, \quad T = 8$						
CML_1	0.002	0.033	0.002	0.023	0.056	0.049	-0.002	0.033	-0.003	0.023	0.056	1.000
CML_0	0.002	0.033	0.002	0.023	0.056		0.038	0.049	0.038	0.038	0.236	
						$n = 250, \quad T = 12$						
CML_1	-0.002	0.025	-0.001	0.017	0.048	0.043	-0.003	0.026	-0.004	0.017	0.064	1.000
CML_0	-0.002	0.025	-0.002	0.017	0.052		0.023	0.034	0.023	0.025	0.164	
						$n = 500, \quad T = 4$						
CML_1	-0.001	0.041	-0.001	0.027	0.043	0.055	-0.001	0.042	-0.000	0.029	0.049	1.000
CML_0	-0.000	0.039	0.001	0.027	0.044		0.078	0.086	0.078	0.078	0.566	
						$n = 500, \quad T = 8$						
CML_1	0.001	0.024	0.001	0.016	0.065	0.037	-0.003	0.022	-0.003	0.015	0.046	1.000
CML_0	0.001	0.023	0.000	0.016	0.054		0.037	0.043	0.037	0.037	0.410	
						$n = 500, \quad T = 12$						
CML_1	-0.000	0.018	-0.000	0.012	0.045	0.041	-0.002	0.018	-0.002	0.012	0.063	1.000
CML_0	-0.000	0.018	-0.001	0.012	0.047		0.025	0.031	0.025	0.025	0.322	
						$n = 1000, \quad T = 4$						
CML_1	0.001	0.030	0.000	0.020	0.053	0.051	-0.000	0.028	-0.000	0.018	0.052	1.000
CML_0	0.001	0.028	0.001	0.018	0.050		0.078	0.082	0.078	0.078	0.860	
						$n = 1000, \quad T = 8$						
CML_1	0.000	0.016	0.000	0.011	0.052	0.051	-0.003	0.017	-0.004	0.012	0.050	1.000
CML_0	0.000	0.016	0.001	0.011	0.047		0.037	0.040	0.036	0.036	0.669	
						$n = 1000, \quad T = 12$						
CML_1	-0.000	0.013	-0.000	0.009	0.048	0.045	-0.002	0.013	-0.002	0.008	0.057	1.000
CML_0	-0.000	0.013	-0.000	0.009	0.047		0.025	0.027	0.025	0.025	0.550	

benchmark design: Table 3 refers to a normally distributed covariate depending on the time-varying covariate z_{it}, Table 4 concerns the case of time persistence in

Table 2: Simulation results from Experiment 2: CML estimator, $\mu = -1$, $\psi = 0$, $\rho = 0$, normally distributed covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 100$, $T = 4$						
CML_1	-0.062	0.238	-0.019	0.130	0.048	0.059	-0.093	0.267	-0.057	0.143	0.062	0.601
CML_0	-0.045	0.221	-0.008	0.122	0.053		0.030	0.218	0.053	0.147	0.103	
						$n = 100$, $T = 8$						
CML_1	-0.013	0.094	-0.008	0.064	0.043	0.052	-0.020	0.105	-0.015	0.072	0.064	0.995
CML_0	-0.010	0.092	-0.006	0.062	0.042		0.034	0.101	0.035	0.074	0.093	
						$n = 100$, $T = 12$						
CML_1	-0.010	0.074	-0.007	0.050	0.060	0.062	-0.008	0.078	-0.004	0.052	0.062	1.000
CML_0	-0.009	0.073	-0.005	0.051	0.055		0.025	0.078	0.028	0.053	0.095	
						$n = 250$, $T = 4$						
CML_1	-0.026	0.130	-0.014	0.081	0.053	0.046	-0.025	0.133	-0.012	0.081	0.056	0.952
CML_0	-0.021	0.124	-0.008	0.078	0.055		0.074	0.138	0.086	0.107	0.158	
						$n = 250$, $T = 8$						
CML_1	-0.005	0.060	-0.002	0.041	0.045	0.054	-0.009	0.066	-0.006	0.043	0.054	1.000
CML_0	-0.004	0.060	-0.001	0.040	0.047		0.040	0.072	0.041	0.052	0.149	
						$n = 250$, $T = 12$						
CML_1	-0.004	0.046	-0.003	0.029	0.070	0.061	-0.002	0.046	-0.002	0.032	0.050	1.000
CML_0	-0.003	0.046	-0.001	0.029	0.067		0.030	0.052	0.030	0.037	0.124	
						$n = 500$, $T = 4$						
CML_1	-0.013	0.087	-0.007	0.055	0.043	0.038	-0.019	0.092	-0.016	0.061	0.049	0.999
CML_0	-0.010	0.084	-0.004	0.055	0.043		0.081	0.113	0.086	0.088	0.225	
						$n = 500$, $T = 8$						
CML_1	-0.004	0.043	-0.003	0.029	0.057	0.037	-0.005	0.044	-0.003	0.029	0.054	1.000
CML_0	-0.004	0.043	-0.002	0.028	0.054		0.042	0.059	0.044	0.047	0.203	
						$n = 500$, $T = 12$						
CML_1	-0.003	0.031	-0.002	0.020	0.039	0.055	-0.003	0.033	-0.002	0.022	0.048	1.000
CML_0	-0.003	0.031	-0.001	0.020	0.040		0.029	0.042	0.029	0.031	0.158	
						$n = 1000$, $T = 4$						
CML_1	-0.005	0.061	-0.002	0.039	0.049	0.058	-0.006	0.060	-0.005	0.040	0.038	1.000
CML_0	-0.004	0.058	-0.001	0.039	0.038		0.088	0.103	0.090	0.091	0.382	
						$n = 1000$, $T = 8$						
CML_1	-0.002	0.029	-0.002	0.019	0.045	0.040	-0.002	0.031	-0.002	0.021	0.047	1.000
CML_0	-0.002	0.029	-0.003	0.019	0.049		0.046	0.054	0.045	0.045	0.353	
						$n = 1000$, $T = 12$						
CML_1	-0.001	0.023	-0.001	0.015	0.054	0.063	-0.000	0.023	-0.000	0.015	0.057	1.000
CML_0	-0.001	0.023	-0.000	0.015	0.053		0.031	0.038	0.031	0.031	0.295	

x_{it} formulated through AR(1) errors, while Table 5 refers to a binary covariate. These scenarios allow us to investigate the properties of the CML estimator when the assumption formulated by equation (13) does not hold and the model formulated in Theorem 1 just embeds a linear approximation of this equation. The results for $\psi = 0.5$ in Table 3 mirror closely those in Table 2, whereas the presence of AR(1) errors in the conditional mean for x_{it} does not seem to hamper the ability of CML_1 to produce consistent estimates of μ. When the covariate is binary, instead, the bias of the CML_1 estimator of μ is almost always negligible. Regarding efficiency, the RMSE and MAE are slightly higher for μ, although they decrease with both n and T (see Table 5).

Table 6 summarizes the simulation results based on the PCML estimator for the design where there is state dependence in the response variable, that is $\delta = 1$. Regarding the estimation of μ, the Table depicts a similar situation depicted by Table 2, whereas the estimator of δ exhibits a certain bias when T is small. It is worth recalling that the PCML estimator is consistent only with $\delta = 0$, while it provides a good approximation of δ when $\delta \neq 0$. Moreover, the performance of the PCML estimator may be especially sensitive to the degree of state dependence in the generated samples. A high value of δ leads to a reduction of the actual sample size and represents a large deviation from the approximating point $\delta = 0$. Nevertheless, Bartolucci and Nigro (2012) showed that the bias and root-mean square error of PCML estimator of δ in the dynamic logit model decrease at a rate close to $\sqrt{n}$ and as T grows also for δ moving away from 0. Regarding the static formulation, Experiments 7 to 9 explore the violations of assumption (13). The results for both μ and δ also suggest that the proposed formulation provides a good approximation when time-varying covariates enter the conditional mean of x_{it} or the distribution of x_{it} does not belong to the exponential family (see Tables 7 and 9). The estimator PCML_1 performs well even when x_{it} has some persistence formulated through AR(1) distributed errors (see Table 8), even though the estimator of δ has a nonnegligible bias when $T = 4$.

Table 3: Simulation results from Experiment 3: CML estimator, $\mu = -1$, $\psi = 0.5$, $\rho = 0$, normally distributed covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
$n = 100$, $T = 4$												
CML_1	-0.076	0.279	-0.040	0.146	0.052	0.050	-0.102	0.302	-0.060	0.158	0.069	0.478
CML_0	-0.057	0.259	-0.021	0.136	0.059		0.014	0.232	0.044	0.147	0.102	
$n = 100$, $T = 8$												
CML_1	-0.014	0.102	-0.011	0.068	0.044	0.065	-0.015	0.111	-0.007	0.073	0.050	0.970
CML_0	-0.010	0.100	-0.008	0.066	0.044		0.030	0.108	0.036	0.076	0.087	
$n = 100$, $T = 12$												
CML_1	-0.010	0.078	-0.008	0.054	0.058	0.063	-0.005	0.081	0.001	0.055	0.058	0.997
CML_0	-0.008	0.077	-0.005	0.054	0.058		0.022	0.081	0.027	0.057	0.090	
$n = 250$, $T = 4$												
CML_1	-0.036	0.149	-0.016	0.091	0.058	0.054	-0.021	0.146	-0.002	0.085	0.075	0.858
CML_0	-0.028	0.143	-0.011	0.087	0.057		0.063	0.144	0.076	0.105	0.149	
$n = 250$, $T = 8$												
CML_1	-0.006	0.061	-0.003	0.041	0.033	0.052	-0.004	0.067	-0.002	0.044	0.056	1.000
CML_0	-0.005	0.060	-0.001	0.041	0.032		0.035	0.073	0.037	0.053	0.121	
$n = 250$, $T = 12$												
CML_1	-0.005	0.049	-0.004	0.031	0.063	0.058	-0.001	0.048	0.001	0.034	0.042	1.000
CML_0	-0.004	0.049	-0.003	0.032	0.062		0.025	0.053	0.025	0.037	0.092	
$n = 500$, $T = 4$												
CML_1	-0.016	0.093	-0.010	0.061	0.048	0.052	-0.011	0.095	-0.006	0.063	0.053	0.988
CML_0	-0.012	0.090	-0.005	0.057	0.049		0.069	0.110	0.072	0.081	0.184	
$n = 500$, $T = 8$												
CML_1	-0.003	0.045	-0.002	0.030	0.053	0.042	-0.000	0.047	0.003	0.032	0.055	1.000
CML_0	-0.002	0.045	-0.000	0.030	0.053		0.038	0.059	0.041	0.045	0.167	
$n = 500$, $T = 12$												
CML_1	-0.002	0.033	-0.002	0.021	0.049	0.058	-0.001	0.035	-0.001	0.023	0.056	1.000
CML_0	-0.002	0.033	-0.001	0.021	0.049		0.025	0.042	0.025	0.030	0.140	
$n = 1000$, $T = 4$												
CML_1	-0.006	0.064	-0.004	0.043	0.037	0.062	-0.003	0.067	0.001	0.046	0.053	1.000
CML_0	-0.004	0.062	-0.002	0.042	0.040		0.074	0.096	0.076	0.079	0.278	
$n = 1000$, $T = 8$												
CML_1	-0.002	0.031	-0.002	0.021	0.041	0.049	0.003	0.034	0.003	0.022	0.065	1.000
CML_0	-0.002	0.031	-0.002	0.020	0.041		0.041	0.052	0.041	0.042	0.275	
$n = 1000$, $T = 12$												
CML_1	-0.001	0.024	-0.000	0.017	0.046	0.048	0.003	0.025	0.003	0.017	0.057	1.000
CML_0	-0.001	0.024	-0.001	0.017	0.048		0.028	0.037	0.029	0.029	0.240	

Table 4: Simulation results from Experiment 4: CML estimator, $\mu = -1$, $\psi = 0$, $\rho = 0.25$, normally distributed covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
$n = 100$, $T = 4$												
CML_1	-0.082	0.242	-0.051	0.127	0.058	0.071	-0.081	0.259	-0.056	0.147	0.052	0.584
CML_0	-0.058	0.221	-0.028	0.122	0.059		0.048	0.211	0.073	0.143	0.110	
$n = 100$, $T = 8$												
CML_1	-0.012	0.093	-0.001	0.060	0.046	0.051	-0.018	0.101	-0.010	0.067	0.039	0.996
CML_0	-0.009	0.091	0.002	0.060	0.049		0.037	0.100	0.043	0.072	0.083	
$n = 100$, $T = 12$												
CML_1	-0.006	0.071	-0.003	0.047	0.053	0.053	-0.007	0.077	-0.006	0.049	0.060	1.000
CML_0	-0.004	0.070	-0.002	0.048	0.055		0.026	0.078	0.028	0.054	0.091	
$n = 250$, $T = 4$												
CML_1	-0.021	0.130	-0.010	0.085	0.046	0.051	-0.030	0.134	-0.020	0.085	0.052	0.950
CML_0	-0.016	0.126	-0.005	0.085	0.046		0.082	0.144	0.093	0.110	0.170	
$n = 250$, $T = 8$												
CML_1	-0.005	0.060	-0.002	0.040	0.054	0.048	-0.005	0.062	-0.003	0.040	0.050	1.000
CML_0	-0.004	0.060	0.001	0.039	0.053		0.043	0.073	0.044	0.054	0.137	
$n = 250$, $T = 12$												
CML_1	-0.005	0.046	-0.004	0.030	0.052	0.061	-0.003	0.048	-0.003	0.032	0.058	1.000
CML_0	-0.004	0.046	-0.003	0.030	0.054		0.028	0.053	0.029	0.038	0.126	
$n = 500$, $T = 4$												
CML_1	-0.008	0.088	0.004	0.058	0.053	0.057	-0.014	0.091	-0.009	0.058	0.054	0.999
CML_0	-0.005	0.085	0.002	0.057	0.052		0.091	0.122	0.097	0.099	0.248	
$n = 500$, $T = 8$												
CML_1	-0.003	0.041	-0.003	0.027	0.047	0.057	-0.003	0.045	-0.002	0.030	0.062	1.000
CML_0	-0.002	0.041	-0.002	0.027	0.050		0.045	0.062	0.046	0.047	0.234	
$n = 500$, $T = 12$												
CML_1	-0.002	0.031	-0.001	0.020	0.056	0.054	-0.002	0.034	-0.001	0.023	0.054	1.000
CML_0	-0.002	0.031	-0.001	0.021	0.053		0.028	0.043	0.029	0.031	0.170	
$n = 1000$, $T = 4$												
CML_1	-0.005	0.059	-0.003	0.041	0.044	0.052	-0.010	0.067	-0.008	0.045	0.057	1.000
CML_0	-0.003	0.057	-0.002	0.042	0.041		0.094	0.112	0.096	0.097	0.423	
$n = 1000$, $T = 8$												
CML_1	-0.001	0.030	0.000	0.021	0.054	0.040	-0.001	0.030	-0.000	0.021	0.048	1.000
CML_0	-0.001	0.030	-0.000	0.021	0.051		0.047	0.055	0.046	0.047	0.384	
$n = 1000$, $T = 12$												
CML_1	0.000	0.022	0.001	0.016	0.053	0.052	-0.000	0.022	0.000	0.015	0.043	1.000
CML_0	0.000	0.022	0.001	0.015	0.058		0.030	0.037	0.030	0.030	0.288	

Table 5: Simulation results from Experiment 5: CML estimator, $\mu = -1$, $\psi = 0$, $\rho = 0$, binary covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 100$, $T = 4$						
CML_1	-0.047	0.488	-0.011	0.307	0.057	0.054	-0.075	0.478	-0.058	0.298	0.050	0.583
CML_0	-0.040	0.450	-0.005	0.291	0.052		0.247	0.484	0.262	0.339	0.123	
						$n = 100$, $T = 8$						
CML_1	-0.023	0.228	-0.020	0.146	0.049	0.051	-0.018	0.238	-0.017	0.156	0.057	0.987
CML_0	-0.022	0.222	-0.018	0.138	0.051		0.126	0.256	0.127	0.177	0.103	
						$n = 100$, $T = 12$						
CML_1	-0.014	0.169	-0.012	0.114	0.054	0.060	-0.008	0.176	-0.008	0.113	0.057	1.000
CML_0	-0.013	0.168	-0.013	0.112	0.058		0.082	0.189	0.083	0.129	0.086	
						$n = 250$, $T = 4$						
CML_1	-0.019	0.282	-0.017	0.191	0.051	0.039	-0.006	0.277	-0.007	0.187	0.044	0.934
CML_0	-0.012	0.261	-0.009	0.174	0.051		0.301	0.391	0.303	0.306	0.226	
						$n = 250$, $T = 8$						
CML_1	-0.003	0.143	-0.003	0.092	0.055	0.051	-0.010	0.150	-0.005	0.100	0.069	1.000
CML_0	-0.001	0.141	0.002	0.092	0.052		0.128	0.192	0.132	0.143	0.177	
						$n = 250$, $T = 12$						
CML_1	-0.006	0.105	-0.006	0.066	0.044	0.045	-0.010	0.109	-0.008	0.076	0.038	1.000
CML_0	-0.006	0.104	-0.007	0.065	0.041		0.078	0.131	0.082	0.095	0.116	
						$n = 500$, $T = 4$						
CML_1	-0.013	0.188	-0.009	0.121	0.043	0.054	-0.025	0.201	-0.019	0.128	0.061	0.999
CML_0	-0.013	0.180	-0.013	0.120	0.044		0.286	0.338	0.292	0.292	0.389	
						$n = 500$, $T = 8$						
CML_1	-0.001	0.103	-0.002	0.069	0.052	0.046	-0.006	0.100	-0.004	0.072	0.047	1.000
CML_0	-0.000	0.102	0.001	0.066	0.055		0.133	0.164	0.135	0.136	0.290	
						$n = 500$, $T = 12$						
CML_1	-0.003	0.075	-0.003	0.052	0.045	0.039	-0.008	0.081	-0.010	0.052	0.068	1.000
CML_0	-0.003	0.075	-0.004	0.053	0.044		0.080	0.112	0.079	0.084	0.181	
						$n = 1000$, $T = 4$						
CML_1	-0.006	0.133	-0.007	0.090	0.039	0.053	-0.006	0.140	-0.009	0.094	0.041	1.000
CML_0	-0.005	0.127	-0.002	0.086	0.050		0.300	0.326	0.295	0.295	0.666	
						$n = 1000$, $T = 8$						
CML_1	-0.003	0.069	-0.002	0.046	0.041	0.053	-0.007	0.071	-0.003	0.048	0.053	1.000
CML_0	-0.003	0.068	-0.001	0.045	0.042		0.130	0.147	0.134	0.134	0.494	
						$n = 1000$, $T = 12$						
CML_1	-0.002	0.052	0.000	0.036	0.041	0.043	-0.005	0.057	-0.004	0.039	0.066	1.000
CML_0	-0.002	0.052	-0.000	0.036	0.039		0.084	0.100	0.083	0.083	0.352	

Table 6: Simulation results from Experiment 6: dynamic logit model, PCML estimator, $\mu = -1$, $\delta = 1$, $\psi = 0$, $\rho = 0$, normally distributed covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 500$, $T = 4$						
PCML_1	-0.012	0.186	0.013	0.116	0.056	0.063	-0.024	0.210	0.002	0.132	0.062	0.529
PCML_0	-0.013	0.169	0.000	0.105	0.060		0.133	0.216	0.155	0.173	0.214	
						$n = 500$, $T = 8$						
PCML_1	-0.004	0.050	-0.002	0.034	0.046	0.053	-0.001	0.057	0.002	0.038	0.053	1.000
PCML_0	-0.003	0.049	-0.001	0.033	0.047		0.024	0.059	0.025	0.042	0.092	
						$n = 500$, $T = 12$						
PCML_1	-0.003	0.035	-0.002	0.024	0.039	0.062	-0.002	0.040	-0.003	0.028	0.045	1.000
PCML_0	-0.002	0.035	-0.001	0.024	0.039		-0.001	0.038	-0.001	0.027	0.041	
						$n = 1000$, $T = 4$						
PCML_1	0.021	0.116	0.029	0.078	0.059	0.057	0.024	0.133	0.034	0.093	0.078	0.824
PCML_0	0.010	0.107	0.018	0.076	0.056		0.163	0.197	0.174	0.175	0.380	
						$n = 1000$, $T = 8$						
PCML_1	-0.003	0.036	-0.002	0.023	0.056	0.056	0.004	0.040	0.006	0.029	0.053	1.000
PCML_0	-0.002	0.036	-0.002	0.023	0.058		0.027	0.046	0.030	0.033	0.128	
						$n = 1000$, $T = 12$						
PCML_1	-0.001	0.025	-0.001	0.017	0.053	0.065	0.002	0.027	0.001	0.019	0.049	1.000
PCML_0	-0.000	0.025	-0.000	0.017	0.054		0.003	0.025	0.002	0.018	0.044	
Estimation of δ												
						$n = 500$, $T = 4$						
PCML_1	0.143	0.590	0.108	0.365	0.056	0.063	0.092	0.599	0.082	0.412	0.044	0.529
PCML_0	0.129	0.571	0.109	0.363	0.055		0.188	0.590	0.161	0.385	0.048	
						$n = 500$, $T = 8$						
PCML_1	0.013	0.128	0.010	0.087	0.047	0.053	0.004	0.155	0.005	0.103	0.064	1.000
PCML_0	0.013	0.128	0.010	0.085	0.050		-0.126	0.191	-0.125	0.138	0.166	
						$n = 500$, $T = 12$						
PCML_1	0.006	0.090	0.007	0.062	0.051	0.062	-0.001	0.096	-0.004	0.067	0.036	1.000
PCML_0	0.006	0.089	0.006	0.063	0.049		-0.069	0.116	-0.070	0.084	0.101	
						$n = 1000$, $T = 4$						
PCML_1	0.086	0.386	0.064	0.247	0.057	0.057	0.074	0.404	0.062	0.275	0.054	0.824
PCML_0	0.083	0.383	0.059	0.243	0.054		0.168	0.415	0.155	0.275	0.063	
						$n = 1000$, $T = 8$						
PCML_1	0.013	0.094	0.008	0.066	0.057	0.056	0.009	0.102	0.009	0.068	0.049	1.000
PCML_0	0.013	0.094	0.009	0.066	0.057		-0.120	0.154	-0.120	0.122	0.247	
						$n = 1000$, $T = 12$						
PCML_1	0.002	0.063	0.005	0.043	0.047	0.065	0.002	0.071	0.005	0.048	0.049	1.000
PCML_0	0.002	0.063	0.004	0.043	0.045		-0.066	0.095	-0.063	0.067	0.173	

Table 7: Simulation results from Experiment 7: dynamic logit model, PCML estimator, $\mu = -1$, $\delta = 1$, $\psi = 0.5$, $\rho = 0$, normally distributed covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 500, \quad T = 4$						
PCML$_1$	-0.072	0.263	-0.038	0.141	0.056	0.070	-0.012	0.252	0.017	0.159	0.084	0.328
PCML$_0$	-0.063	0.242	-0.031	0.134	0.059		0.118	0.245	0.145	0.184	0.201	
						$n = 500, \quad T = 8$						
PCML$_1$	-0.004	0.053	-0.002	0.035	0.043	0.055	0.004	0.062	0.009	0.039	0.064	1.000
PCML$_0$	-0.003	0.053	-0.002	0.034	0.042		0.025	0.063	0.027	0.045	0.096	
						$n = 500, \quad T = 12$						
PCML$_1$	-0.002	0.037	-0.001	0.025	0.047	0.065	-0.001	0.039	-0.001	0.026	0.045	1.000
PCML$_0$	-0.001	0.037	-0.001	0.025	0.047		0.002	0.038	0.003	0.025	0.044	
						$n = 1000, \quad T = 4$						
PCML$_1$	-0.011	0.138	0.005	0.085	0.046	0.062	0.035	0.168	0.046	0.111	0.095	0.613
PCML$_0$	-0.008	0.129	0.008	0.085	0.046		0.149	0.204	0.160	0.165	0.285	
						$n = 1000, \quad T = 8$						
PCML$_1$	-0.005	0.037	-0.005	0.025	0.036	0.077	0.004	0.040	0.006	0.028	0.044	1.000
PCML$_0$	-0.005	0.037	-0.004	0.025	0.037		0.024	0.045	0.027	0.033	0.088	
						$n = 1000, \quad T = 12$						
PCML$_1$	-0.002	0.026	-0.001	0.017	0.035	0.107	0.002	0.029	0.002	0.020	0.053	1.000
PCML$_0$	-0.001	0.026	-0.001	0.017	0.035		0.005	0.028	0.005	0.019	0.055	
Estimation of δ												
						$n = 500, \quad T = 4$						
PCML$_1$	0.461	0.940	0.382	0.570	0.066	0.070	0.513	1.028	0.447	0.619	0.078	0.328
PCML$_0$	0.446	0.910	0.386	0.565	0.071		0.534	0.999	0.454	0.605	0.077	
						$n = 500, \quad T = 8$						
PCML$_1$	0.011	0.139	0.010	0.089	0.047	0.055	0.050	0.165	0.048	0.104	0.065	1.000
PCML$_0$	0.013	0.138	0.011	0.090	0.048		-0.054	0.160	-0.058	0.113	0.069	
						$n = 500, \quad T = 12$						
PCML$_1$	0.002	0.096	0.002	0.064	0.047	0.065	0.026	0.107	0.022	0.070	0.055	1.000
PCML$_0$	0.002	0.096	0.003	0.064	0.050		-0.023	0.103	-0.024	0.070	0.064	
						$n = 1000, \quad T = 4$						
PCML$_1$	0.392	0.664	0.388	0.446	0.088	0.062	0.459	0.724	0.437	0.489	0.100	0.613
PCML$_0$	0.391	0.657	0.377	0.444	0.092		0.498	0.742	0.481	0.517	0.120	
						$n = 1000, \quad T = 8$						
PCML$_1$	0.004	0.103	0.005	0.068	0.053	0.077	0.053	0.118	0.051	0.077	0.072	1.000
PCML$_0$	0.006	0.102	0.007	0.067	0.058		-0.050	0.112	-0.053	0.074	0.064	
						$n = 1000, \quad T = 12$						
PCML$_1$	0.003	0.066	0.002	0.044	0.043	0.107	0.019	0.076	0.019	0.053	0.055	1.000
PCML$_0$	0.003	0.066	0.002	0.045	0.040		-0.028	0.077	-0.029	0.056	0.069	

Table 8: Simulation results from Experiment 8: dynamic logit model, PCML estimator, $\mu = -1$, $\delta = 1$, $\psi = 0$, $\rho = 0.25$, normally distributed covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 500$, $T = 4$						
PCML$_1$	-0.037	0.183	-0.016	0.113	0.049	0.065	-0.040	0.224	-0.015	0.135	0.057	0.490
PCML$_0$	-0.026	0.168	-0.007	0.109	0.047		0.097	0.204	0.117	0.153	0.157	
						$n = 500$, $T = 8$						
PCML$_1$	-0.005	0.049	-0.005	0.033	0.043	0.062	-0.004	0.057	-0.000	0.038	0.052	1.000
PCML$_0$	-0.004	0.049	-0.004	0.033	0.042		0.019	0.056	0.022	0.039	0.079	
						$n = 500$, $T = 12$						
PCML$_1$	-0.002	0.035	-0.000	0.023	0.062	0.072	-0.002	0.038	-0.002	0.026	0.050	1.000
PCML$_0$	-0.001	0.035	0.000	0.023	0.060		-0.000	0.037	-0.001	0.025	0.051	
						$n = 1000$, $T = 4$						
PCML$_1$	-0.006	0.123	0.001	0.079	0.058	0.055	0.014	0.142	0.027	0.100	0.059	0.817
PCML$_0$	-0.004	0.115	0.006	0.078	0.053		0.138	0.183	0.145	0.154	0.312	
						$n = 1000$, $T = 8$						
PCML$_1$	-0.004	0.037	-0.003	0.025	0.050	0.088	0.002	0.038	0.002	0.026	0.044	1.000
PCML$_0$	-0.003	0.036	-0.002	0.025	0.051		0.024	0.043	0.024	0.031	0.108	
						$n = 1000$, $T = 12$						
PCML$_1$	-0.002	0.024	-0.002	0.016	0.048	0.112	0.003	0.028	0.004	0.020	0.055	1.000
PCML$_0$	-0.001	0.024	-0.001	0.016	0.047		0.003	0.027	0.004	0.018	0.059	
Estimation of δ												
						$n = 500$, $T = 4$						
PCML$_1$	0.486	0.881	0.417	0.535	0.090	0.065	0.430	0.907	0.386	0.559	0.073	0.490
PCML$_0$	0.474	0.857	0.417	0.513	0.090		0.433	0.885	0.394	0.552	0.081	
						$n = 500$, $T = 8$						
PCML$_1$	0.012	0.131	0.010	0.087	0.048	0.062	0.020	0.145	0.024	0.102	0.039	1.000
PCML$_0$	0.013	0.131	0.010	0.086	0.050		-0.089	0.162	-0.085	0.108	0.094	
						$n = 500$, $T = 12$						
PCML$_1$	0.004	0.091	0.001	0.060	0.053	0.072	0.008	0.099	0.010	0.067	0.044	1.000
PCML$_0$	0.004	0.091	0.001	0.060	0.053		-0.046	0.105	-0.044	0.071	0.071	
						$n = 1000$, $T = 4$						
PCML$_1$	0.394	0.620	0.360	0.392	0.108	0.055	0.366	0.635	0.368	0.442	0.093	0.817
PCML$_0$	0.395	0.617	0.362	0.400	0.111		0.387	0.637	0.391	0.447	0.104	
						$n = 1000$, $T = 8$						
PCML$_1$	0.014	0.092	0.010	0.060	0.052	0.088	0.016	0.106	0.017	0.070	0.057	1.000
PCML$_0$	0.015	0.092	0.012	0.061	0.050		-0.091	0.135	-0.091	0.098	0.145	
						$n = 1000$, $T = 12$						
PCML$_1$	0.003	0.062	0.003	0.042	0.049	0.112	0.007	0.073	0.010	0.050	0.068	1.000
PCML$_0$	0.003	0.062	0.003	0.042	0.048		-0.045	0.084	-0.045	0.058	0.100	

Table 9: Simulation results from Experiment 9: dynamic logit model, PCML estimator, $\mu = -1$, $\delta = 1$, $\psi = 0$, $\rho = 0$, binary covariate

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 500,\quad T = 4$						
PCML$_1$	-0.061	0.384	-0.037	0.244	0.058	0.048	0.011	0.411	0.023	0.271	0.059	0.584
PCML$_0$	-0.046	0.341	-0.029	0.216	0.058		0.418	0.535	0.430	0.436	0.269	
						$n = 500,\quad T = 8$						
PCML$_1$	-0.007	0.115	-0.004	0.080	0.046	0.058	0.010	0.119	0.010	0.076	0.046	1.000
PCML$_0$	-0.007	0.114	-0.004	0.081	0.047		0.051	0.123	0.049	0.081	0.073	
						$n = 500,\quad T = 12$						
PCML$_1$	-0.003	0.079	-0.003	0.054	0.039	0.078	0.006	0.082	0.007	0.053	0.035	1.000
PCML$_0$	-0.001	0.080	-0.000	0.055	0.042		-0.036	0.087	-0.036	0.059	0.059	
						$n = 1000,\quad T = 4$						
PCML$_1$	-0.018	0.258	-0.017	0.178	0.039	0.034	0.036	0.277	0.052	0.188	0.051	0.892
PCML$_0$	-0.018	0.226	-0.011	0.160	0.041		0.433	0.487	0.439	0.439	0.509	
						$n = 1000,\quad T = 8$						
PCML$_1$	-0.010	0.082	-0.007	0.055	0.058	0.077	0.007	0.087	0.009	0.057	0.053	1.000
PCML$_0$	-0.010	0.082	-0.006	0.055	0.057		0.047	0.094	0.045	0.064	0.097	
						$n = 1000,\quad T = 12$						
PCML$_1$	-0.003	0.058	-0.004	0.040	0.049	0.112	0.006	0.062	0.006	0.043	0.045	1.000
PCML$_0$	-0.001	0.058	-0.001	0.039	0.050		-0.036	0.068	-0.036	0.047	0.100	
Estimation of δ												
						$n = 500,\quad T = 4$						
PCML$_1$	0.320	0.606	0.316	0.405	0.079	0.048	0.307	0.592	0.318	0.409	0.075	0.584
PCML$_0$	0.311	0.595	0.310	0.395	0.076		0.359	0.609	0.363	0.434	0.092	
						$n = 500,\quad T = 8$						
PCML$_1$	0.009	0.113	0.007	0.075	0.046	0.058	-0.002	0.116	-0.003	0.085	0.045	1.000
PCML$_0$	0.008	0.112	0.008	0.075	0.046		0.004	0.112	0.004	0.081	0.051	
						$n = 500,\quad T = 12$						
PCML$_1$	0.002	0.081	0.004	0.054	0.062	0.078	-0.004	0.082	-0.003	0.054	0.053	1.000
PCML$_0$	0.001	0.081	0.003	0.053	0.061		0.004	0.079	0.007	0.054	0.046	
						$n = 1000,\quad T = 4$						
PCML$_1$	0.280	0.454	0.278	0.310	0.115	0.034	0.294	0.462	0.303	0.329	0.118	0.892
PCML$_0$	0.276	0.449	0.271	0.312	0.113		0.350	0.495	0.352	0.369	0.146	
						$n = 1000,\quad T = 8$						
PCML$_1$	0.001	0.079	-0.000	0.052	0.046	0.077	-0.002	0.082	-0.003	0.053	0.053	1.000
PCML$_0$	0.001	0.079	-0.002	0.052	0.046		0.004	0.078	0.004	0.051	0.053	
						$n = 1000,\quad T = 12$						
PCML$_1$	0.002	0.057	0.003	0.038	0.055	0.112	-0.009	0.061	-0.012	0.042	0.069	1.000
PCML$_0$	0.002	0.057	0.002	0.038	0.057		-0.001	0.058	-0.006	0.040	0.064	

5.3. Comparison with alternative estimators

As already discussed in Section 1, the proposed approach represents a competing alternative to the Wooldridge (2000)'s method in two main respects: first, dealing with feedback amounts to simply adding leads of the predetermined covariates to the linear index rather than specifying a comprehensive joint model, which has to be specified and implemented on a case-wise basis; secondly, unobserved heterogeneity is treated nonparametrically and is allowed to be correlated with the predetermined covariates, as opposed to the correlated random-effects approach, adopted by Wooldridge, where individual specific effects can only be functions of strictly exogenous covariates.

More in detail, following a Mundlak (1978)-type approach, a specification for the individual effects is assumed in Wooldridge (2000), which is based on

$$\begin{aligned} c_i &= y_{i1}\varpi_1 + x_{i1}\varpi_2 + s_i\varpi_3 + c_i^*, \\ c_i^* &\sim N(0,\sigma_c^2), \\ \xi_i &= \lambda c_i, \end{aligned}$$

for $i = 1,\dots,n$, where $s_i = (1/T)\sum_{t=1}^T z_{it}$. The model parameters $\boldsymbol{\psi} = (\boldsymbol{\theta}', \boldsymbol{\varpi}', \sigma_c, \lambda)'$, where $\boldsymbol{\varpi} = (\varpi_1, \varpi_2, \varpi_3)'$, are estimated by maximizing the log-likelihood function

$$\ell(\boldsymbol{\psi}) = \sum_{i=1}^n \log \int \prod_{t=1}^T p(y_{it}|x_{it}, z_{it}, c_i) f(x_{it}|y_{i,t-1}, z_{it}, c_i) \frac{1}{\sigma_c}\phi\left(\frac{c_i^*}{\sigma_c}\right) \mathrm{d}c_i^*$$

where y_{it} and x_{it} are assumed to follow the same design as in Experiment 5, which includes state dependence, $f(x_{it}|y_{i,t-1}, z_{it}, c_i)$ is the density of x_{it} conditional on $y_{i,t-1}$, z_{it}, and c_i, and $\phi(\cdot)$ is the standard normal density function.

We propose a comparison of the performance of the PCML estimator for model (15) with the random-effects ML estimator illustrated above. This comparison is based on two scenarios. In the first one, the individual intercepts are generated as standard normal random variables independent of both the strictly exogenous and the predetermined covariate. In the second one, the individual effects are generated as in (22) and are therefore correlated with the predetermined covariate. In this case, the assumption for the Mundlak-type correction imposed by Wooldridge (2000) is violated. These last two experiments are denoted as Experiment 10 and Experiment 11, respectively.

Tables 10 and 11 summarize the results of the additional simulation study based on Experiments 10 and 11, which is limited to the scenarios with $\mu = -1$, $\delta = 1$, $\psi = 0$, and $\eta = 0, -1$. In Experiment 10, the biases for μ and δ obtained by PCML and ML and the RMSE and MAE attain the same order of magnitude with $T = 8, 12$. In Experiment 11, the bias for μ obtained by ML is somewhat larger than the one obtained by PCML with $\eta = 0$, whereas, with $\eta = -1$, the magnitude of the bias of the ML and PCML estimators are overall rather similar. Regarding δ, the bias of the ML estimator is higher with $\eta = 0$, while it decreases as T grows when $\eta = -1$.

6. Empirical application

We apply the proposed formulation to the problem of estimating the labor supply of married women as a function of young children in the family, where the presence of children can be predetermined because past labor market participation events may affect present fertility decisions. Our application is closely related to the analyses performed in the literature on feedback effects (Chamberlain, 1984; Carrasco, 2001; Arellano and Carrasco, 2003; Mosconi and Seri, 2006; Michaud and Tatsiramos, 2011).

The empirical analysis is based on a sample drawn from the PSID, that consists of $n = 1,908$ married women between 19 and 59 years of age in 1980, followed for $T = 7$ time occasions, from 1979 to 1985. We specify logit models for the probability of being employed at time t, conditional on the lagged employment status, as well as the number of children of a certain age in the family, namely the number of kids between 0 and 2 years old, between 3 and 5, and between 6 and 17. We also include the husband's income, the woman's

Table 10: Simulation results from Experiment 10: PCML and ML estimators, $\mu = -1$, $\delta = 1$, $\psi = 0$, normally distributed covariate, independent individual effects.

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 500, \quad T = 4$						
PCML	-0.046	0.202	-0.026	0.118	0.050	0.060	-0.032	0.240	0.000	0.141	0.058	0.500
ML	-0.008	0.084	-0.011	0.053	0.993		-0.001	0.092	-0.006	0.056	0.998	
						$n = 500, \quad T = 8$						
PCML	-0.004	0.051	-0.000	0.032	0.056	0.070	0.001	0.059	0.002	0.039	0.055	1.000
ML	0.003	0.069	-0.006	0.027	1.000		0.016	0.086	-0.004	0.036	1.000	
						$n = 500, \quad T = 12$						
PCML	-0.003	0.035	-0.002	0.024	0.050	0.071	0.001	0.040	0.000	0.026	0.061	1.000
ML	-0.009	0.039	-0.011	0.023	0.999		-0.004	0.054	-0.011	0.024	0.999	
						$n = 1000, \quad T = 4$						
PCML	-0.015	0.129	-0.008	0.081	0.047	0.045	0.010	0.146	0.022	0.098	0.060	0.814
ML	-0.006	0.069	-0.013	0.036	0.997		-0.009	0.058	-0.007	0.037	0.997	
						$n = 1000, \quad T = 8$						
PCML	-0.002	0.036	-0.002	0.024	0.052	0.081	0.003	0.040	0.005	0.029	0.046	1.000
ML	-0.005	0.040	-0.007	0.019	0.999		-0.002	0.055	-0.010	0.022	1.000	
						$n = 1000, \quad T = 12$						
PCML	-0.001	0.025	-0.000	0.017	0.051	0.106	0.002	0.027	0.001	0.019	0.044	1.000
ML	-0.009	0.032	-0.011	0.017	1.000		-0.009	0.038	-0.011	0.018	0.999	
Estimation of δ												
						$n = 500, \quad T = 4$						
PCML	0.331	0.759	0.330	0.483	0.045	0.060	0.372	0.873	0.351	0.572	0.062	0.500
ML	-0.053	0.210	-0.053	0.151	0.949		-0.094	0.261	-0.088	0.164	0.922	
						$n = 500, \quad T = 8$						
PCML	0.006	0.139	0.001	0.096	0.056	0.070	-0.011	0.151	-0.015	0.098	0.052	1.000
ML	0.004	0.109	0.006	0.075	0.988		-0.054	0.112	-0.052	0.077	0.994	
						$n = 500, \quad T = 12$						
PCML	0.003	0.093	0.003	0.065	0.050	0.071	-0.003	0.101	-0.001	0.073	0.041	1.000
ML	0.050	0.092	0.050	0.062	0.997		-0.015	0.081	-0.011	0.057	0.994	
						$n = 1000, \quad T = 4$						
PCML	0.026	0.293	0.024	0.187	0.047	0.054	0.325	0.598	0.345	0.421	0.079	0.814
ML	-0.077	0.152	-0.078	0.107	0.966		-0.104	0.195	-0.106	0.135	0.975	
						$n = 1000, \quad T = 8$						
PCML	0.340	0.594	0.302	0.385	0.106	0.045	-0.011	0.104	-0.014	0.069	0.049	1.000
ML	-0.047	0.160	-0.051	0.114	0.974		-0.051	0.122	-0.057	0.065	0.997	
						$n = 1000, \quad T = 12$						
PCML	-0.001	0.025	-0.000	0.017	0.051	0.106	-0.006	0.075	-0.006	0.049	0.062	1.000
ML	-0.009	0.032	-0.011	0.017	1.000		-0.015	0.058	-0.015	0.039	0.997	

Table 11: Simulation results from Experiment 11: PCML and ML estimators, $\mu = -1$, $\delta = 1$, $\psi = 0$, normally distributed covariate, individual effects generated as in (22).

	$\eta = 0$						$\eta = -1$					
	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$	Mean bias	RMSE	Median bias	MAE	t-test	t-test $H_0 : \nu = 0$
Estimation of μ												
						$n = 500$, $T = 4$						
PCML	0.008	0.184	0.028	0.125	0.034	0.045	-0.009	0.277	0.016	0.182	0.056	0.991
ML	-0.029	0.078	-0.027	0.050	0.992		-0.023	0.095	-0.024	0.054	0.973	
						$n = 500$, $T = 8$						
PCML	-0.007	0.062	-0.008	0.041	0.049	0.054	0.023	0.085	0.026	0.058	0.062	1.000
ML	-0.034	0.060	-0.036	0.040	0.998		-0.015	0.092	-0.031	0.041	0.990	
						$n = 500$, $T = 12$						
PCML	-0.001	0.045	0.000	0.029	0.055	0.063	0.015	0.061	0.016	0.041	0.070	1.000
ML	-0.031	0.050	-0.031	0.035	1.000		-0.009	0.095	-0.028	0.039	0.995	
						$n = 1000$, $T = 4$						
PCML	0.028	0.130	0.038	0.090	0.051	0.054	0.030	0.189	0.042	0.132	0.073	1.000
ML	-0.014	0.093	-0.027	0.041	1.000		-0.016	0.071	-0.019	0.037	0.994	
						$n = 1000$, $T = 8$						
PCML	-0.003	0.046	-0.002	0.032	0.054	0.062	0.025	0.062	0.028	0.045	0.076	1.000
ML	-0.032	0.046	-0.032	0.033	1.000		-0.009	0.090	-0.026	0.032	0.999	
						$n = 1000$, $T = 12$						
PCML	-0.002	0.032	-0.001	0.022	0.051	0.068	0.016	0.042	0.015	0.029	0.058	1.000
ML	-0.031	0.040	-0.031	0.032	1.000		-0.027	0.046	-0.027	0.029	0.999	
Estimation of δ												
						$n = 500$, $T = 4$						
PCML	0.043	0.430	0.022	0.290	0.054	0.045	0.065	0.571	0.062	0.379	0.056	0.991
ML	-0.080	0.206	-0.088	0.149	0.947		-0.162	0.301	-0.168	0.211	0.903	
						$n = 500$, $T = 8$						
PCML	0.002	0.121	0.007	0.085	0.055	0.054	-0.027	0.158	-0.028	0.106	0.056	1.000
ML	0.064	0.109	0.066	0.080	0.995		-0.034	0.114	-0.038	0.081	0.987	
						$n = 500$, $T = 12$						
PCML	0.003	0.085	0.008	0.057	0.056	0.063	-0.017	0.110	-0.014	0.072	0.052	1.000
ML	0.124	0.142	0.126	0.126	0.999		0.022	0.090	0.016	0.057	0.995	
						$n = 1000$, $T = 4$						
PCML	0.026	0.293	0.024	0.187	0.047	0.054	0.045	0.371	0.043	0.253	0.042	1.000
ML	-0.077	0.152	-0.078	0.107	0.966		-0.145	0.229	-0.145	0.164	0.965	
						$n = 1000$, $T = 8$						
PCML	0.009	0.083	0.008	0.056	0.051	0.062	-0.029	0.116	-0.029	0.079	0.061	1.000
ML	0.077	0.098	0.078	0.078	0.996		-0.028	0.088	-0.031	0.058	0.997	
						$n = 1000$, $T = 12$						
PCML	0.002	0.059	0.002	0.039	0.058	0.068	-0.015	0.077	-0.013	0.052	0.049	1.000
ML	0.124	0.133	0.124	0.124	0.999		0.018	0.061	0.018	0.040	0.998	

age, and time fixed-effects. Another relevant covariate is the level of education but it exhibits no time variation in the sample considered. It is therefore not included as its effect on the response probability is not identified.

We specify a static logit model for female labor supply as in (6), which rules out feedback effects, and as the proposed formulation in (14), which admits violations of S. Similarly, we also specify a dynamic logit model under noncausality as in (7) and a model that allows for violations of S' as in (15). In order to allow for departures from strict exogeneity and noncausality, we consider in both models the number of kids between 0 and 2 years old, 3 and 5, and between 6 and 17 in the family and the husband's income as predetermined. We estimate the two static logit models by CML and the two dynamic logit models by PCML. It is worth noticing that, for this case, the strategy proposed by Wooldridge (2000) illustrated in Section 5.3 would require the specification of a five-equation model, where the parameters of the main equation are jointly estimated along with those in the four equations specified for the predetermined covariates.

Table 12 reports the estimation results for the static logit model, along with the estimates of the average partial effects. It emerges that the only predetermined variable seems to be that related to the presence of children between 3 and 5 years old in the household. The associated coefficient is slightly smaller when we allow for violations of the strict exogeneity assumption and the corresponding average partial effects becomes not statistically significant.

Consistently with related empirical results on female labor supply, the results reported in Table 13 show that labor force participation is highly persistent, as the estimate of the state dependent parameter is close to 1.5 whit both PCML_0 and PCML_1, meaning that women employed in $t-1$ have a probability of working in t that is, on average, 10 percentage points higher than for women not working in $t-1$. With this specification, the effects of the number of children and husband's income is no longer statistically significant and so are the corresponding partial effects. In addition, none of the leads included in the model specification seem to capture departures form noncausality.

Table 12: Female labor force participation: logit model

	Model parameters		Average partial effects	
	CML$_0$	CML$_1$	CML$_0$	CML$_1$
# Children 0-2$_t$	-1.074***	-0.952***	-0.075***	-0.066**
	(0.137)	(0.138)	(0.028)	(0.028)
# Children 3-5$_t$	-0.750***	-0.655***	-0.052*	-0.046
	(0.148)	(0.151)	(0.030)	(0.031)
# Children 6-17$_t$	-0.217	-0.282**	-0.015	-0.020
	(0.136)	(0.144)	(0.028)	(0.030)
Husband income$_t$	-0.019***	-0.021***	-0.001	-0.001
	(0.006)	(0.006)	(0.001)	(0.001)
Age$_t$	0.328	-0.043	-0.045	0.055
	(1.390)	(1.458)	(0.068)	(0.071)
Age squared$_t$	-0.138	-0.107		
	(0.192)	(0.199)		
# Children 0-2$_{t+1}$		-0.230		
		(0.157)		
# Children 3-5$_{t+1}$		-0.290**		
		(0.152)		
# Children 6-17$_{t+1}$		0.078		
		(0.133)		
Husband income$_{t+1}$		0.005		
		(0.005)		

Notes: standard errors in square brackets. *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.10. Estimates are based on $T = 6$ time occasions for each woman. Source: PSID 1979-1985.

7. Conclusions

We propose a novel model formulation for binary logit panel data models that accounts for feedback effects from the the outcome variable on the future values of the covariates. Our proposal is particularly well suited for short panels with a large number of cross-section units, typically provided by rotated or strongly unbalanced continuous surveys, often employed in microeconomic applications. Within this setting, the proposed formulation lends itself to a fixed-effects estimation approach based on conditional inference.

The proposed model formulation yields two main advantages compared to the few available alternatives: (i) it does not require the specification of a parametric model for the predetermined explanatory variables; (ii) it has a simple formulation and, in practice, it allows us to include a large number of predeter-

Table 13: Female labor force participation: dynamic logit model

	Model parameters		Average partial effects	
	CML$_0$	CML$_1$	CML$_0$	CML$_1$
Employed$_{t-1}$	1.475*** (0.149)	1.468*** (0.151)	0.107*** (0.032)	0.106*** (0.032)
# Children 0-2$_t$	-0.718*** (0.165)	-0.707*** (0.176)	-0.046 (0.031)	-0.045 (0.033)
# Children 3-5$_t$	-0.486*** (0.164)	-0.476*** (0.180)	-0.031 (0.031)	-0.030 (0.035)
# Children 6-17$_t$	-0.167 (0.141)	-0.261 (0.166)	-0.011 (0.027)	-0.017 (0.032)
Husband income$_t$	-0.014** (0.007)	-0.016** (0.007)	-0.001 (0.001)	-0.001 (0.001)
Age$_t$	1.783 (1.560)	1.245 (1.617)	0.014 (0.068)	0.008 (0.071)
Age squared$_t$	-0.224 (0.209)	-0.161 (0.215)		
# Children 0-2$_{t+1}$		0.024 (0.201)		
# Children 3-5$_{t+1}$		0.009 (0.190)		
# Children 6-17$_{t+1}$		0.202 (0.173)		
Husband income$_{t+1}$		0.006 (0.005)		

Notes: standard errors in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value $<$ 0.10. Estimates are based on $T = 5$ time occasions for each woman. Source: PSID 1979-1985.

mined covariates, either discrete or continuous.

From our simulation results, it emerges that the CML and PCML estimators have good finite-sample performance when applied to the proposed models in presence of substantial departures from noncausality. Also the finite-sample bias of the estimators is negligible even when the conditions for the exact logit model formulation proposed in this paper are violated. Furthermore, we show that an alternative correlated random-effects estimator has comparable finite-sample properties for $T \geq 8$, while the PCML outperforms the ML estimator one with a reduced number of time periods.

Finally, the logit model here proposed is fairly easy to estimate using available software. The CML and PCML estimators of the proposed models can be implemented using the package `cquad` (Bartolucci and Pigini, 2017).

Acknowledgements

We acknowledge one anonymous Reviewers and the Associate Editor for insightful comments that helped us improve the paper. Francesco Bartolucci acknowledges the financial support from the University of Perugia that funded the project (Ricerca di Base, 2018) entitled "Partial effects in econometric models for binary longitudinal data based on quadratic exponential distributions".

References

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

Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598–606.

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

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of Econometrics 68, 29–51.

Arellano, M., Carrasco, R., 2003. Binary choice panel data models with predetermined variables. Journal of Econometrics 115, 125–157.

Barndorff-Nielsen, O., 1978. Information and Exponential Families in Statistical Theory. John Wiley & Sons.

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

Bartolucci, F., 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., 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., 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.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143.

Carrasco, R., 2001. Binary choice with binary endogenous regressors in panel data. Journal of Business & Economic Statistics 19, 385–394.

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

Chamberlain, G., 1980. Analysis of covariance with qualitative data. The Review of Economic Studies 47, 225–238.

Chamberlain, G., 1982. The general equivalence of Granger and Sims causality. Econometrica 50, 569–581.

Chamberlain, G., 1984. Panel data. Handbook of Econometrics 2, 1247–1318.

Chamberlain, G., 1985. Heterogeneity, omitted variable bias, and duration dependence, in: Heckman, J.J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data. Cambridge University Press: Cambridge, pp. 3–38.

Cox, D., 1972. The analysis of multivariate binary data. Applied Statistics 21, 113–120.

Dumitrescu, E.I., Hurlin, C., 2012. Testing for Granger non-causality in heterogeneous panels. Economic Modelling 29, 1450–1460.

Emirmahmutoglu, F., Kose, N., 2011. Testing for Granger causality in heterogeneous mixed panels. Economic Modelling 28, 870–876.

Fernandez-Val, I., 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit model. Journal of Econometrics 150, 71–85.

Florens, J.P., Mouchart, M., 1982. A note on noncausality. Econometrica 50, 583–591.

Granger, C.W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

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

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

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Heckman, J.J., 1981. Heterogeneity and state dependence. Structural Analysis of Discrete Data with Econometric Applications MIT Press: Cambridge MA. Manski CF, McFadden (eds).

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

Honoré, B.E., Lewbel, A., 2002. Semiparametric binary choice panel data models without strictly exogeneous regressors. Econometrica 70, 2053–2063.

Hsiao, C., 2015. Analysis of Panel Data. 3rd ed., Cambridge University Press, New York.

Michaud, P.C., Tatsiramos, K., 2011. Fertility and female employment dynamics in Europe: The effect of using alternative econometric modeling assumptions. Journal of Applied Econometrics 26, 641–668.

Mosconi, R., Seri, R., 2006. Non-causality in bivariate binary time series. Journal of Econometrics 132, 379–407.

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

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

Sims, C.A., 1972. Money, income, and causality. The American Economic Review 62, 540–552.

Wooldridge, J.M., 2000. A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. Economics Letters 20, 245–250.

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. The MIT press, Cambridge, MA.

Appendix A. Proof of the equivalence theorem

Theorem 2. G *and* S' *are equivalent conditions.*

Proof. G may be reformulated as

$$\frac{p(x_{i,t+1}, c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t})}{p(c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t})} = \frac{p(x_{i,t+1}, c_i, \boldsymbol{x}_{i,1:t})}{p(c_i, \boldsymbol{x}_{i,1:t})}, \quad t = 1, \ldots, T-1,$$

for all i. Exchanging the denominator at lhs with the numerator at rhs, the previous equality becomes

$$p(\boldsymbol{y}_{i,1:t}|c_i, \boldsymbol{x}_{i,1:t+1}) = p(\boldsymbol{y}_{i,1:t}|c_i, \boldsymbol{x}_{i,1:t}), \quad t = 1, \ldots, T-1,$$

which, by marginalization, implies that

$$p(\boldsymbol{y}_{i,1:s}|c_i, \boldsymbol{x}_{i,1:t+1}) = p(\boldsymbol{y}_{i,1:s}|c_i, \boldsymbol{x}_{i,1:t}), \quad t = 1, \ldots, T-1,\ s = 1, \ldots, t.$$

Therefore, we have

$$p(y_{is}|c_i, \boldsymbol{x}_{i,1:t+1}, \boldsymbol{y}_{i,1:s-1}) = p(y_{is}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:s-1}), \quad t = 1, \ldots, T-1, s = 1, \ldots, t.$$

Finally, by recursively using the previous expression for a fixed s and for t from $T-1$ to s we obtain condition S' as defined in (2). Similarly, S' implies that

$$p(\boldsymbol{x}_{i,t+1:T}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t}) = p(\boldsymbol{x}_{i,t+1:T}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t-1}), \quad t = 1, \ldots, T-1,$$

for all i and implies

$$p(x_{i,s+1}|c_i, \boldsymbol{x}_{i,1:s}, \boldsymbol{y}_{i,1:t}) = p(x_{i,s+1}|c_i, \boldsymbol{x}_{i,1:s}, \boldsymbol{y}_{i,1:t-1}), \quad t = 1, \ldots, T-1, s = 1, \ldots, T-1,$$

which, in turn, leads to condition (3) and then G. □

Theorem 3. G *implies* S.

Proof. Proceeding as in the proof of Theorem 2, G implies that

$$p(y_{is}|c_i, \boldsymbol{x}_{i,1:t+1}) = p(y_{is}|c_i, \boldsymbol{x}_{i,1:t}), \quad t = 1, \dots, T-1, \; s = 1, \dots, t.$$

By recursively using the previous expression for a fixed s and for t from $T-1$ to s, we obtain condition (1). □