Econometrica, Vol. 78, No. 2 (March, 2010), 719–733

A DYNAMIC MODEL FOR BINARY PANEL DATA WITH UNOBSERVED HETEROGENEITY ADMITTING A $\sqrt{n}$-CONSISTENT CONDITIONAL ESTIMATOR

BY FRANCESCO BARTOLUCCI AND VALENTINA NIGRO[1]

A model for binary panel data is introduced which allows for state dependence and unobserved heterogeneity beyond the effect of available covariates. The model is of quadratic exponential type and its structure closely resembles that of the dynamic logit model. However, it has the advantage of being easily estimable via conditional likelihood with at least two observations (further to an initial observation) and even in the presence of time dummies among the regressors.

KEYWORDS: Longitudinal data, quadratic exponential distribution, state dependence.

1. INTRODUCTION

BINARY PANEL DATA ARE USUALLY ANALYZED by using a dynamic logit or probit model which includes, among the explanatory variables, the lags of the response variable and has individual-specific intercepts; see Arellano and Honoré (2001) and Hsiao (2005), among others. These models allow us to disentangle the true *state dependence* (i.e., how the experience of an event in the past can influence the occurrence of the same event in the future) from the propensity to experience a certain outcome in all periods, when the latter depends on unobservable factors (see Heckman (1981a, 1981b)). State dependence arises in many economic contexts, such as job decision, investment choice, and brand choice, and can determine different policy implications. The parameters of main interest in these models are typically those for the covariates and the true state dependence, which are referred to as *structural parameters*. The individual-specific intercepts are referred to as *incidental parameters*; they are of interest only in certain situations, such as when we need to obtain marginal effects and predictions.

In this paper, we introduce a model for binary panel data which closely resembles the dynamic logit model, and, as such, allows for state dependence and unobserved heterogeneity between subjects, beyond the effect of the available covariates. The model is a version of the quadratic exponential model (Cox (1972)) with covariates in which (i) the first-order effects depend on the covari-

[1]We thank a co-editor and three anonymous referees for helpful suggestions and insightful comments. We are also grateful to Franco Peracchi and Frank Vella for their comments and suggestions. Francesco Bartolucci acknowledges financial support from the Einaudi Institute for Economics and Finance (EIEF), Rome. Most of the article was developed during the period Valentina Nigro spent at the University of Rome "Tor Vergata" and is part of her Ph.D. dissertation.

 DOI: 10.3982/ECTA7531

ates and on an individual-specific parameter for the unobserved heterogeneity, and (ii) the second-order effects are equal to a common parameter when they are referred to pairs of consecutive response variables and to 0 otherwise. We show that this parameter has the same interpretation that it has in the dynamic logit model in terms of *log-odds ratio*, a measure of association between binary variables which is well known in the statistical literature on categorical data analysis (Agresti (2002, Chap. 8)). For the proposed model, we also provide a justification as a latent index model in which the systematic component depends on expectation about future outcomes, beyond the covariates and the lags of the response variable, and the stochastic component has a standard logistic distribution.

An important feature of the proposed model is that, as for the static logit model, the incidental parameters can be eliminated by conditioning on sufficient statistics for these parameters, which correspond to the sums of the response variables at individual level. Using a terminology derived from Rasch (1961), these statistics will be referred to as *total scores*. The resulting conditional likelihood allows us to identify the structural parameters for the covariates and the state dependence with at least two observations (further to an initial observation). The estimator of the structural parameters based on the maximization of this function is $\sqrt{n}$-consistent; moreover, it is simpler to compute than the estimator of Honoré and Kyriazidou (2000) and may be used even in the presence of time dummies. On the basis of a simulation study, the results of which are reported in the Supplemental Material file (Bartolucci and Nigro (2010)), we also notice that the estimator has good finite-sample properties in terms of both bias and efficiency.

The paper is organized as follows. In the next section, we briefly review the dynamic logit model for binary panel data. The proposed model is described in Section 3, where we also show that the total scores are sufficient statistics for its incidental parameters. Identification of the structural parameters and the conditional maximum likelihood estimator of these parameters is illustrated in Section 4.

2. DYNAMIC LOGIT MODEL FOR BINARY PANEL DATA

In the following discussion, we first review the dynamic logit model for binary panel data; then we discuss conditional inference and related inferential methods on its structural parameters.

2.1. *Basic Assumptions*

Let y_{it} be a binary response variable equal to 1 if subject i ($i = 1, \dots, n$) makes a certain choice at time t ($t = 1, \dots, T$) and equal to 0 otherwise; also let $\mathbf{x}_{it}$ be a corresponding vector of strictly exogenous covariates. The standard fixed-effects approach for binary panel data assumes that

$$
\begin{aligned}
(1)\qquad & y_{it} = 1\{y_{it}^* \geq 0\}, \\
& y_{it}^* = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma + \varepsilon_{it}, \quad i = 1, \dots, n,\ t = 1, \dots, T,
\end{aligned}
$$

where $1\{\cdot\}$ is the indicator function and y_{it}^* is a latent variable which may be interpreted as utility (or propensity) of the choice. Moreover, the zero-mean random variables ε_{it} represent error terms. Of primary interest are the vector of parameters for the covariates, $\boldsymbol{\beta}$, and the parameter that measures the state dependence effect, γ. These are the structural parameters which are collected in the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \gamma)'$. The individual-specific intercepts α_i are instead the incidental parameters.

The error terms ε_{it} are typically assumed to be independent and identically distributed conditionally on the covariates and the individual-specific parameters, and assumed to have a standard logistic distribution. The conditional distribution of y_{it} given α_i, $\mathbf{X}_i = (\,\mathbf{x}_{i1} \quad \cdots \quad \mathbf{x}_{iT}\,)$ and $y_{i0}, \dots, y_{i,t-1}$ can then be expressed as

$$
\begin{aligned}
(2)\qquad p(y_{it}|\alpha_i, \mathbf{X}_i, y_{i0}, \dots, y_{i,t-1}) &= p(y_{it}|\alpha_i, \mathbf{x}_{it}, y_{i,t-1}) \\
&= \frac{\exp[y_{it}(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]}{1 + \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)}
\end{aligned}
$$

for $i = 1, \dots, n$ and $t = 1, \dots, T$. This is a dynamic logit formulation which implies the following conditional distribution of the overall vector of response variables $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$ given α_i, $\mathbf{X}_i$ and y_{i0}:

$$
(3)\qquad p(\mathbf{y}_i|\alpha_i, \mathbf{X}_i, y_{i0}) = \frac{\exp\left(y_{i+}\alpha_i + \sum_t y_{it}\mathbf{x}_{it}'\boldsymbol{\beta} + y_{i*}\gamma\right)}{\prod_t [1 + \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]},
$$

where $y_{i+} = \sum_t y_{it}$ and $y_{i*} = \sum_t y_{i,t-1}y_{it}$, with the sum $\sum_t$ and the product $\prod_t$ ranging over $t = 1, \dots, T$. The statistic y_{i+} is referred to as the total score of subject i.

For what follows, it is important to note that

$$
\log \frac{p(y_{it} = 0|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 0)\,p(y_{it} = 1|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 1)}{p(y_{it} = 0|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 1)\,p(y_{it} = 1|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 0)} = \gamma
$$

for $i = 1, \dots, n$ and $t = 1, \dots, T$. Thus, the parameter γ for the state dependence corresponds to the conditional log-odds ratio between $(y_{i,t-1}, y_{it})$ for every i and t.

2.2. *Conditional Inference*

As mentioned in Section 1, an effective approach to estimate the model illustrated above is based on the maximization of the conditional likelihood given suitable sufficient statistics.

For the static version of the model, in which the parameter γ is equal to 0, we have that $\mathbf{y}_i$ is conditionally independent of α_i given y_{i0}, $\mathbf{X}_i$, and the total score y_{i+}, and then $p(\mathbf{y}_i|\alpha_i, \mathbf{X}_i, y_{i+}) = p(\mathbf{y}_i|\mathbf{X}_i, y_{i+})$. The likelihood based on this conditional probability allows us to identify $\boldsymbol{\beta}$ for $T \geq 2$; by maximizing this likelihood we also obtain a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$. Even though referred to a simpler context, this result goes back to Rasch (1961) and was developed by Andersen (1970). See also Magnac (2004), who characterized other situations in which the total scores are sufficient statistics for the individual-specific intercepts.

Among the first authors to deal with the conditional approach for the dynamic logit model (γ is unconstrained) were Cox (1958) and Chamberlain (1985). In particular, the latter noticed that when $T = 3$ and the covariates are omitted from the model, $p(\mathbf{y}_i|\alpha_i, y_{i0}, y_{i1} + y_{i2} = 1, y_{i3})$ does not depend on α_i for every y_{i0} and y_{i3}. On the basis of this conditional distribution, it is therefore possible to construct a likelihood function which depends on the response configurations of only certain subjects (those such that $y_{i1} + y_{i2} = 1$), and which allows us to identify and consistently estimate the parameter γ.

The approach of Chamberlain (1985) was extended by Honoré and Kyriazidou (2000) to the case where, as in (2), the model includes exogenous covariates. In particular, when these covariates are continuous, they proposed to estimate the vector $\boldsymbol{\theta}$ of structural parameters by maximizing a weighted conditional log-likelihood with weights depending on the individual covariates through a kernel function which must be defined in advance.

Although the weighted conditional approach of Honoré and Kyriazidou (2000) is of great interest, their results about identification and consistency are based on certain assumptions on the support of the covariates which rule out, for instance, time dummies. Moreover, the approach requires careful choice of the kernel function and of its bandwidth, since these choices affect the performance of their estimator. Furthermore, the estimator is consistent as $n \to \infty$, but its rate of convergence to the true parameter value is slower than $\sqrt{n}$, unless only discrete covariates are present. See also Magnac (2004) and Honoré and Tamer (2006).

Even though it is not strictly related to the conditional approach, it is worth mentioning that a recent line of research investigated dynamic discrete choice models with fixed-effects proposing bias corrected estimators (see Hahn and Newey (2004), Carro (2007)). Although these estimators are only consistent when the number of time periods goes to infinity, they have a reduced order of the bias without increasing the asymptotic variance. Monte Carlo simulations have shown their good finite-sample performance in comparison to the estimator of Honoré and Kyriazidou (2000) even with not very long panels (e.g., seven time periods).

3. PROPOSED MODEL FOR BINARY PANEL DATA

In this section, we introduce a quadratic exponential model for binary panel data and we discuss its main features in comparison to the dynamic logit model.

3.1. *Basic Assumptions*

We assume that

$$p(\mathbf{y}_i|\alpha_i, \mathbf{X}_i, y_{i0}) = \frac{\exp\left[y_{i+}\alpha_i + \sum_t y_{it}\mathbf{x}_{it}'\boldsymbol{\beta}_1 + y_{iT}(\phi + \mathbf{x}_{iT}'\boldsymbol{\beta}_2) + y_{i*}\gamma\right]}{\sum_{\mathbf{z}} \exp\left[z_+\alpha_i + \sum_t z_t\mathbf{x}_{it}'\boldsymbol{\beta}_1 + z_T(\phi + \mathbf{x}_{iT}'\boldsymbol{\beta}_2) + z_{i*}\gamma\right]}, \tag{4}$$

where the sum $\sum_{\mathbf{z}}$ ranges over all possible binary response vectors $\mathbf{z} = (z_1, \dots, z_T)'$; moreover, $z_+ = \sum_t z_t$ and $z_{i*} = y_{i0}z_1 + \sum_{t>1} z_{t-1}z_t$. The denominator does not depend on $\mathbf{y}_i$; it is simply a *normalizing constant* that we denote by $\mu(\alpha_i, \mathbf{X}_i, y_{i0})$. The model can be viewed as a version of the quadratic exponential model of Cox (1972) with covariates in which the first-order effect for y_{it} is equal to $\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_1$ (to which we add $\phi + \mathbf{x}_{it}'\boldsymbol{\beta}_2$ when $t = T$) and the second-order effect for (y_{is}, y_{it}) is equal to γ when $t = s + 1$ and equal to 0 otherwise. The need for a different parametrization of the first-order effect when $t = T$ and $t < T$ will be clarified below.

It is worth noting that the expression for the probability of $\mathbf{y}_i$ given in (4) closely resembles that given in (3) which results from the dynamic logit model. From some simple algebra, we also obtain that

$$\log \frac{p(y_{it} = 0|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 0)\, p(y_{it} = 1|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 1)}{p(y_{it} = 0|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 1)\, p(y_{it} = 1|\alpha_i, \mathbf{X}_i, y_{i,t-1} = 0)} = \gamma$$

for every i and t. Then, under the proposed quadratic exponential model, γ has the same interpretation that it has under the dynamic logit model, that is, log-odds ratio between each pair of consecutive response variables. Not surprisingly, the dynamic logit model coincides with the proposed model in the absence of state dependence ($\gamma = 0$).[2]

[2]It is also possible to show that, up to a correction term, expression (4) is an approximation of that in (3) obtained by a first-order Taylor expansion around $\alpha_i = 0$, $\boldsymbol{\beta} = \mathbf{0}$, and $\gamma = 0$.

The main difference with respect to the dynamic logit is in the resulting conditional distribution of y_{it} given the available covariates $\mathbf{X}_i$ and $y_{i0}, \dots, y_{i,t-1}$. In fact, (4) implies that

$$(5)\qquad p(y_{it}|\alpha_i, \mathbf{X}_i, y_{i0}, \dots, y_{i,t-1}) = \frac{\exp\{y_{it}[\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_1 + y_{i,t-1}\gamma + e_t^*(\alpha_i, \mathbf{X}_i)]\}}{1 + \exp[\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_1 + y_{i,t-1}\gamma + e_t^*(\alpha_i, \mathbf{X}_i)]},$$

where, for $t < T$,

$$(6)\qquad e_t^*(\alpha_i, \mathbf{X}_i) = \log \frac{1 + \exp[\alpha_i + \mathbf{x}_{i,t+1}'\boldsymbol{\beta}_1 + e_{t+1}^*(\alpha_i, \mathbf{X}_i) + \gamma]}{1 + \exp[\alpha_i + \mathbf{x}_{i,t+1}'\boldsymbol{\beta}_1 + e_{t+1}^*(\alpha_i, \mathbf{X}_i)]} = \log \frac{p(y_{i,t+1} = 0|\alpha_i, \mathbf{X}_i, y_{it} = 0)}{p(y_{i,t+1} = 0|\alpha_i, \mathbf{X}_i, y_{it} = 1)}$$

and

$$(7)\qquad e_T^*(\alpha_i, \mathbf{X}_i) = \phi + \mathbf{x}_{iT}'\boldsymbol{\beta}_2.$$

Then, for $t = T$, the proposed model is equivalent to a dynamic logit model with a suitable parametrization. The interpretation of this correction term will be discussed in detail in Section 3.2. For the moment, it is important to note that the conditional probability depends on present and future covariates, meaning that these covariates are not strictly exogenous (see Wooldridge (2001, Sec. 15.8.2)). The relation between the covariates and the feedback of the response variables vanishes when $\gamma = 0$. Consider also that, for $t < T$, the same Taylor expansion mentioned in footnote 2 leads to $e_t^*(\alpha_i, \mathbf{X}_i) \approx 0.5\gamma$. Under this approximation, $p(y_{it}|\alpha_i, \mathbf{X}_i, y_{i0}, \dots, y_{i,t-1})$ does not depend on the future covariates and these covariates can be considered strictly exogenous in an approximate sense.

In the simpler case without covariates, the conditional probability of y_{it} becomes

$$p(y_{it}|\alpha_i, y_{i0}, \dots, y_{i,t-1}) = \frac{\exp\{y_{it}[\alpha_i + y_{i,t-1}\gamma + e_t^*(\alpha_i)]\}}{1 + \exp[\alpha_i + y_{i,t-1}\gamma + e_t^*(\alpha_i)]}, \qquad t = 1, \dots, T-1,$$

whereas, for the last period, we have the logistic parametrization

$$p(y_{iT}|\alpha_i, y_{i0}, \dots, y_{i,T-1}) = \frac{\exp[y_{iT}(\alpha_i + y_{i,T-1}\gamma)]}{1 + \exp(\alpha_i + y_{i,T-1}\gamma)},$$

where

$$e_t^*(\alpha_i) = \log \frac{p(y_{i,t+1} = 0|\alpha_i, y_{it} = 0)}{p(y_{i,t+1} = 0|\alpha_i, y_{it} = 1)},$$

which is 0 only in the absence of state dependence.

Finally, we have to clarify that the possibility to use quadratic exponential models for panel data is already known in the statistical literature; see Diggle, Heagerty, Liang, and Zeger (2002) and Molenberghs and Verbeke (2004). However, the parametrization adopted in this type of literature, which is different from the one we propose, is sometimes criticized for lack of a simple interpretation. In contrast, for our parametrization, we provide a justification as a latent index model.

3.2. *Model Justification and Related Issues*

Expression (5) implies that the proposed model is equivalent to the latent index model

$$(8) \qquad y_{it} = 1\{y^*_{it} \geq 0\}, \quad y^*_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}_1 + y_{i,t-1}\gamma + e^*_t(\alpha_i, \mathbf{X}_i) + \varepsilon_{it},$$

where the error terms ε_{it} are independent and have standard logistic distribution. Assumption (8) is similar to assumption (1) on which the dynamic logit model is based, the main difference being in the correction term $e^*_t(\alpha_i, \mathbf{X}_i)$. As is clear from (6), this term can be interpreted as a measure of the effect of the present choice y_{it} on the expected utility (or propensity) at the next occasion $(t+1)$. In the presence of positive state dependence ($\gamma > 0$), this correction term is positive, since making the choice today has a positive impact on the expected utility. Also note that the different definition of $e^*_t(\alpha_i, \mathbf{X}_i)$ for $t < T$ and $t = T$ (compare equations (6) and (7)) is motivated by considering that $e^*_T(\alpha_i, \mathbf{X}_i)$ has an unspecified form, because it would depend on future covariates not in $\mathbf{X}_i$; then we assume this term to be equal to a linear form of the covariates $\mathbf{x}_{iT}$, in a way similar to that suggested by Heckman (1981c) to deal with the initial condition problem.

As suggested by a referee, it is possible to justify formulation (8), which involves the correction term for the expectation, on the basis of an extension of the job search model described by Hyslop (1999). The latter is based on the maximization of a discounted utility and relies on a budget constraint in which search costs are considered only for subjects who did not participate in the labor market in the previous year. In our extension, subjects who decide to not participate in the current year save an amount of these costs for the next year, but benefit from the amounts previously saved according to the same rule. The reservation wage is then modified so that the decision to participate depends on future expectation about the participation state, beyond the past state. This motivates the introduction of the correction term $e^*_t(\alpha_i, \mathbf{X}_i)$ in (8), which accounts for the difference between the behavior of a subject who has a budget constraint including expectation about future search costs and a subject who has a budget constraint that does not include this expectation.

Two issues that are worth discussing so as to complete the description of the properties of the model are (i) model consistency with respect to marginalizations over a subset of the response variables and (ii) how to avoid assumption (7) on the last correction term.

Assume that (4) holds for the T response variables in $\mathbf{y}_i$. For the subsequence of responses $\mathbf{y}_i^{(T-1)}$, where in general $\mathbf{y}_i^{(t)} = (y_{i1}, \dots, y_{it})'$, we have

$$\begin{aligned} p(\mathbf{y}_i^{(T-1)}|\alpha_i, \mathbf{X}_i, y_{i0}) &= \exp\left[\sum_{t<T} y_{it}(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_1) + \sum_{t<T} y_{i,t-1}y_{it}\gamma\right] \\ &\quad \times [1 + \exp(\phi + \mathbf{x}_{iT}'\boldsymbol{\delta} + y_{i,T-1}\gamma)]/\mu(\alpha_i, \mathbf{X}_i, y_{i0}) \end{aligned}$$

with $\boldsymbol{\delta} = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$. After some algebra, this expression can be reformulated as

$$(9) \qquad \begin{aligned} p(\mathbf{y}_i^{(T-1)}|\alpha_i, \mathbf{X}_i, y_{i0}) = \frac{\exp\left[\sum_{t<T} y_{it}(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_1) + \sum_{t<T} y_{i,t-1}y_{it}\gamma + y_{i,T-1}e_{T-1}(\alpha_i, \mathbf{X}_i)\right]}{\mu_{T-1}(\alpha_i, \mathbf{X}_i, y_{i0})} \end{aligned}$$

with

$$e_{T-1}(\alpha_i, \mathbf{X}_i) = \log \frac{1 + \exp(\phi + \mathbf{x}_{iT}'\boldsymbol{\delta} + \gamma)}{1 + \exp(\phi + \mathbf{x}_{iT}'\boldsymbol{\delta})}$$

and $\mu_{T-1}(\alpha_i, \mathbf{X}, y_{i0})$ denoting the normalizing constant, which is equal to the sum of the numerator of (9) for all possible configurations of the first $T-1$ response variables. Note that $e_{T-1}(\alpha_i, \mathbf{X}_i)$ has an interpretation similar to the correction term $e_{T-1}^*(\alpha_i, \mathbf{X}_i)$ for the future expectation which is defined above.

When $\gamma = 0$, $e_{T-1}(\alpha_i, \mathbf{X}_i) = 0$ and then $p(\mathbf{y}_i^{(T-1)}|\alpha_i, \mathbf{X}_i, y_{i0}) = p(\mathbf{y}_i^{(T-1)}|\alpha_i, \mathbf{X}_i^{(T-1)}, y_{i0})$ with $\mathbf{X}_i^{(t)} = (\mathbf{x}_{i1} \ \cdots \ \mathbf{x}_{it})$. The latter probability can be expressed as in (4) and model consistency with respect to marginalization exactly holds. In the other cases, this form of consistency approximately holds, in the sense that by substituting $e_{T-1}(\alpha_i, \mathbf{X}_i)$ with its linear approximation, we obtain a distribution $p(\mathbf{y}_i^{(T-1)}|\alpha_i, \mathbf{X}_i^{(T-1)}, y_{i0})$ which can be cast into (4). This argument can be iterated to show that, at least approximately, model consistency holds with respect to marginalizations over an arbitrary number of response variables[3]; in this case, the distribution of interest is $p(\mathbf{y}_i^{(t)}|\alpha_i, \mathbf{X}_i^{(t)}, y_{i0})$ with t smaller than $T-1$.

[3]Simulation results (see the Supplemental Material file) show that, for different values of γ, the bias of the conditional estimator of the structural parameters is negligible and is comparable to that resulting from computing these estimators on the complete data sequence.

Finally, assumption (7) on the last correction term $e_T^*(\alpha_i, \mathbf{X}_i)$ can be avoided by conditioning the joint distribution on the corresponding outcome y_{iT}. This removes this correction term since we have

$$p(y_{i1}, \ldots, y_{iT-1}|\alpha_i, \mathbf{X}_i, y_{i0}, y_{iT}) = \frac{\exp\left[\sum_{t<T} y_{it}\alpha_i + \sum_{t<T} y_{it}\mathbf{x}_{it}'\boldsymbol{\beta}_1 + y_{i*}\gamma\right]}{\mu_{T-1}(\alpha_i, \mathbf{X}_i, y_{i0}, y_{iT})}.$$

This conditional version of the proposed model also has the advantage of being consistent across T. However, it would need at least three observations (beyond the initial one) to make the model parameters identifiable. Moreover, the conditional estimator becomes less efficient with respect to the same estimator applied to the initial model.

3.3. *Conditional Distribution Given the Total Score*

The main advantage of the proposed model with respect to the dynamic logit model is that the total scores y_{i+}, $i = 1, \ldots, n$, represent a set of sufficient statistics for the incidental parameters α_i. This is because, for every i, $\mathbf{y}_i$ is conditionally independent of α_i given $\mathbf{X}_i$, y_{i0}, and y_{i+}.

First of all, note that, under assumption (4),

$$\begin{aligned} p(y_{i+}|\alpha_i, \mathbf{X}_i, y_{i0}) &= \sum_{\mathbf{z}(y_{i+})} p(\mathbf{y}_i = \mathbf{z}|\alpha_i, \mathbf{X}_i, y_{i0}) \\ &= \frac{\exp(y_{i+}\alpha_i)}{\mu(\alpha_i, \mathbf{X}_i, y_{i0})} \sum_{\mathbf{z}(y_{i+})} \exp\left[\sum_t z_t\mathbf{x}_{it}'\boldsymbol{\beta}_1 + z_T(\phi + \mathbf{x}_{iT}'\boldsymbol{\beta}_2) + z_{i*}\gamma\right], \end{aligned}$$

where the sum $\sum_{\mathbf{z}(y_{i+})}$ is restricted to all response configurations $\mathbf{z}$ such that $z_+ = y_{i+}$. After some algebra, the conditional distribution at issue becomes

$$\begin{aligned} p(\mathbf{y}_i|\alpha_i, \mathbf{X}_i, y_{i0}, y_{i+}) &= \frac{p(\mathbf{y}_i|\alpha_i, \mathbf{X}_i, y_{i0})}{p(y_{i+}|\alpha_i, \mathbf{X}_i, y_{i0})} \\ &= \frac{\exp\left[\sum_t y_{it}\mathbf{x}_{it}'\boldsymbol{\beta}_1 + y_T(\phi + \mathbf{x}_{iT}'\boldsymbol{\beta}_2) + y_{i*}\gamma\right]}{\sum_{\mathbf{z}(y_{i+})} \exp\left[\sum_t z_t\mathbf{x}_{it}'\boldsymbol{\beta}_1 + z_T(\phi + \mathbf{x}_{iT}'\boldsymbol{\beta}_2) + z_{i*}\gamma\right]}. \end{aligned} \tag{10}$$

The expression above does not depend on α_i and, therefore, is also denoted by $p(\mathbf{y}_i|\mathbf{X}_i, y_{i0}, y_{i+})$. The same circumstance happens for the elements of $\boldsymbol{\beta}_1$ that correspond to the covariates which are time constant. To make this clearer, consider that we can divide the numerator and the denominator of (10) by $\exp(y_{i+}\mathbf{x}'_{i1}\boldsymbol{\beta}_1)$ and, after rearranging terms, we obtain

$$p(\mathbf{y}_i|\mathbf{X}_i, y_{i0}, y_{i+}) = \frac{\exp\left[\sum_{t>1} y_{it}\mathbf{d}'_{it}\boldsymbol{\beta}_1 + y_{iT}(\phi + \mathbf{x}'_{iT}\boldsymbol{\beta}_2) + y_{i*}\gamma\right]}{\sum_{\mathbf{z}(y_{i+})} \exp\left[\sum_{t>1} z_t\mathbf{d}'_{it}\boldsymbol{\beta}_1 + z_T(\phi + \mathbf{x}'_{iT}\boldsymbol{\beta}_2) + z_{i*}\gamma\right]} \tag{11}$$

with $\mathbf{d}_{it} = \mathbf{x}_{it} - \mathbf{x}_{i1},\ t = 2, \dots, T$. We consequently assume that $\boldsymbol{\beta}_1$ does not include any intercept common to all time occasions and regression parameters for covariates which are time constant; if these parameters are included, they would not be identified. This is typical of other conditional approaches, such as that of Honoré and Kyriazidou (2000), and of fixed-effects approaches in which the individual intercepts are estimated together with the structural parameters. Similarly, $\boldsymbol{\beta}_2$ must not contain any intercept for the last occasion, since this is already included through ϕ.

4. CONDITIONAL INFERENCE ON THE STRUCTURAL PARAMETERS

In the following discussion, we introduce a conditional likelihood based on (11). We also provide formal arguments on the identification of the structural parameters via this function and on the asymptotic properties of the estimator that results from its maximization.

4.1. *Structural Parameters Identification via Conditional Likelihood*

For an observed sample $(\mathbf{X}_i, y_{i0}, \mathbf{y}_i)$, $i = 1, \dots, n$, the conditional likelihood has logarithm

$$\ell(\boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\} \log[p_{\boldsymbol{\theta}}(\mathbf{y}_i|\mathbf{X}_i, y_{i0}, y_{i+})], \tag{12}$$

where the subscript $\boldsymbol{\theta}$ has been added to $p(\cdot|\cdot)$ to indicate that this probability, which is defined in (11), depends on $\boldsymbol{\theta}$. Note that in this case $\boldsymbol{\theta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \phi, \gamma)'$. Also note that the response configurations $\mathbf{y}_i$ with sum 0 or T are removed since these do not contain information on $\boldsymbol{\theta}$.

To obtain a simple expression for the score and the information matrix corresponding to $\ell(\boldsymbol{\theta})$, consider that (11) may be expressed in the canonical exponential family form as

$$p_{\boldsymbol{\theta}}(\mathbf{y}_i|\mathbf{X}_i, y_{i0}, y_{i+}) = \frac{\exp[\mathbf{u}(y_{i0}, \mathbf{y}_i)'\mathbf{A}(\mathbf{X}_i)'\boldsymbol{\theta}]}{\sum_{\mathbf{z}(y_{i+})}\exp[\mathbf{u}(y_{i0}, \mathbf{z})'\mathbf{A}(\mathbf{X}_i)'\boldsymbol{\theta}]},$$

where $\mathbf{u}(y_{i0}, \mathbf{y}_i) = (y_{i2}, \dots, y_{iT}, y_{i*})'$ and

$$\mathbf{A}(\mathbf{X}_i) = \begin{pmatrix} \mathbf{d}_{i2} & \cdots & \mathbf{d}_{i,T-1} & \mathbf{d}_{iT} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{x}_{iT} & \mathbf{0} \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix},$$

with $\mathbf{0}$ denoting a column vector of zeros of suitable dimension. From standard results on exponential family distributions (Barndorff-Nielsen (1978, Chap. 8)), it is easy to obtain

$$\mathbf{s}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\}\mathbf{A}(\mathbf{X}_i)\mathbf{v}_{\boldsymbol{\theta}}(\mathbf{X}_i, y_{i0}, \mathbf{y}_i),$$

$$\mathbf{J}(\boldsymbol{\theta}) = -\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\}\mathbf{A}(\mathbf{X}_i)\mathbf{V}_{\boldsymbol{\theta}}(\mathbf{X}_i, y_{i0}, y_{i+})\mathbf{A}(\mathbf{X}_i)',$$

where

$$\mathbf{v}_{\boldsymbol{\theta}}(\mathbf{X}_i, y_{i0}, \mathbf{y}_i) = \mathbf{u}(y_{i0}, \mathbf{y}_i) - E_{\boldsymbol{\theta}}[\mathbf{u}(y_{i0}, \mathbf{y}_i)|\mathbf{X}_i, y_{i0}, y_{i+}],$$
$$\mathbf{V}_{\boldsymbol{\theta}}(\mathbf{X}_i, y_{i0}, y_{i+}) = V_{\boldsymbol{\theta}}[\mathbf{u}(y_{i0}, \mathbf{y}_i)|\mathbf{X}_i, y_{i0}, y_{i+}].$$

Suppose now that the subjects in the samples are independent of each other with α_i, $\mathbf{X}_i$, y_{i0}, and $\mathbf{y}_i$ drawn, for $i = 1, \dots, n$, from the *true model*

$$f_0(\alpha, \mathbf{X}, y_0, \mathbf{y}) = f_0(\alpha, \mathbf{X}, y_0)p_0(\mathbf{y}|\alpha, \mathbf{X}, y_0), \tag{13}$$

where $f_0(\alpha, \mathbf{X}, y_0)$ denotes the joint distribution of the individual-specific intercept, the covariates $\mathbf{X} = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_T)$, and the initial observation y_0. Furthermore, $p_0(\mathbf{y}|\alpha, \mathbf{X}, y_0)$ denotes the conditional distribution of the response variables under the quadratic exponential model (4) when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, with $\boldsymbol{\theta}_0$ denoting the true value of its structural parameters. Under this assumption, we have that $\widehat{Q}(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta})/n$ converges in probability to $Q_0(\boldsymbol{\theta}) = E_0[\ell(\boldsymbol{\theta})/n] = E_0\{\log[p_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{X}, y_0, y_+)]\}$ for any $\boldsymbol{\theta}$, where $E_0(\cdot)$ denotes the expected value under the true model.

By simple algebra, it is possible to show that the first derivative $\nabla_{\boldsymbol{\theta}}Q(\boldsymbol{\theta})$ is equal to $\mathbf{0}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and that, provided $E_0[\mathbf{A}(X)\mathbf{A}(\mathbf{X})']$ is of full rank, the second derivative matrix $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}Q(\boldsymbol{\theta})$ is always negative definite. This implies that

$Q_0(\boldsymbol{\theta})$ is strictly concave with its only maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and, therefore, the vector of structural parameters is identified.

Note that the regularity condition that $E_0[\mathbf{A}(\mathbf{X})\mathbf{A}(\mathbf{X})']$ is of full rank, necessary to ensure that $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} Q(\boldsymbol{\theta})$ is negative definite, rules out cases of time-constant covariates (see also the discussion in Section 3.3). It is also worth noting that the structural parameters of the model are identified with $T \geq 2$, whereas identification of the structural parameters of the dynamic logit model is only possible when $T \geq 3$ (Chamberlain (1993)). See also the discussion provided by Honoré and Tamer (2006).

4.2. *Conditional Maximum Likelihood Estimator*

The conditional maximum likelihood estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2', \hat{\phi}, \hat{\gamma})'$, is obtained by maximizing the conditional log-likelihood $\ell(\boldsymbol{\theta})$. This maximum may be found by a simple iterative algorithm of Newton–Raphson type. At the hth step, this algorithm updates the estimate of $\boldsymbol{\theta}$ at the previous step, $\boldsymbol{\theta}^{(h-1)}$, as $\boldsymbol{\theta}^{(h)} = \boldsymbol{\theta}^{(h-1)} + \mathbf{J}(\boldsymbol{\theta}^{(h-1)})^{-1}\mathbf{s}(\boldsymbol{\theta}^{(h-1)})$.

Note that the information matrix $\mathbf{J}(\boldsymbol{\theta})$ is always nonnegative definite since it corresponds to the sum of a series of variance–covariance matrices. Provided $E_0[\mathbf{A}(X)\mathbf{A}(\mathbf{X})']$ is of full rank, $\mathbf{J}(\boldsymbol{\theta})$ is also positive definite with probability approaching 1 as $n \to \infty$. Then we can reasonably expect that $\ell(\boldsymbol{\theta})$ is strictly concave and has its unique maximum at $\hat{\boldsymbol{\theta}}$ in most economic applications, where the sample size is usually large. Since we also have that the parameter space is equal to $\mathbb{R}^k$, with k denoting the dimension of $\boldsymbol{\theta}$, the above algorithm is very simple to implement and usually converges in a few steps to $\hat{\boldsymbol{\theta}}$, regardless of the starting value $\boldsymbol{\theta}^{(0)}$.

Under the true model (13), and provided that $E_0[\mathbf{A}(X)\mathbf{A}(\mathbf{X})']$ exists and is of full rank, we have that $\hat{\boldsymbol{\theta}}$ exists, is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$, and has asymptotic Normal distribution as $n \to \infty$. This results may be proved on the basis of standard asymptotic results (cf. Theorems 2.7 and 3.1 of Newey and McFadden (1994)).

From Newey and McFadden (1994, Sec. 4.2), we also derive that the standard errors for the elements of $\hat{\boldsymbol{\theta}}$ can be obtained as the corresponding diagonal elements of $(\hat{\mathbf{J}})^{-1}$ under square root. Note that $\hat{\mathbf{J}}$ is obtained as a by-product from the Newton–Raphson algorithm described above. These standard errors can be used to construct confidence intervals for the parameters and to test hypotheses on them in the usual way.

To study the finite-sample properties of the conditional estimator, we performed a simulation study (for a detailed description, see the Supplemental Material file) that closely follows the one performed by Honoré and Kyriazidou (2000). In particular, we first considered a benchmark design under which samples of different size are generated from the quadratic exponential model (4) for 3 and 7 time occasions, only one covariate generated from a Normal distribution, and different values of γ between 0.25 and 2. As in Honoré and Kyriazidou (2000), we also considered other scenarios based on more sophisticated designs for the regressors. Under each scenario, we generated a suitable number of samples and, for every sample, we computed the proposed conditional estimator, whose property were mainly evaluated in terms of *median bias* and *median absolute error* (MAE). We also computed the corresponding standard errors and obtained confidence intervals with different levels for each structural parameter.

On the basis of the simulation study, we conclude that, for each structural parameter, the bias of the conditional estimator is always negligible (with the exception of the estimator $\hat{\gamma}$ when n is small); this bias tends to increase with γ, to decrease with n, and to decrease very quickly with T. Similarly, we observe that the MAE decreases with n at a rate close to $\sqrt{n}$ and much faster with T. This depends on the fact that the number of observations that contribute to the conditional likelihood increases more than proportionally with T, as an increase of T also determines an increase of the *actual sample size*.[4] Moreover, the MAE of the estimator of each parameter increases with γ. This is mainly due to the fact that when γ is positive, its increase implies a decrease of the actual sample size. The simulation results also show that the confidence intervals based on the conditional estimator attain the nominal level for each parameter. This confirms the validity of the rule to compute standard errors based on the information matrix $\hat{\mathbf{J}}$.

Given the same interpretation of the parameters of the quadratic exponential and the dynamic logit models, it is quite natural to compare the proposed conditional estimator with available estimators of the parameters of the latter model. In particular, the results of our simulation study can be compared with those of Honoré and Kyriazidou (2000). It emerges that our estimator performs better than their estimator in terms of both bias and efficiency. This is mainly due to the fact that the former exploits a larger number of response configurations with respect to the latter. Similarly, our estimator can be compared with the bias corrected estimator proposed by Carro (2007). In this case, we observe that the former performs much better than the latter when the parameter of interest is γ, whereas our estimator performs slightly worse than that of Carro (2007) when the parameters of interest are those in $\boldsymbol{\beta}_1$. However, when considering these conclusions, one must be conscious that the results compared here derive from simulation studies performed under different, although very similar, models.

REFERENCES

AGRESTI, A. (2002): *Categorical Data Analysis* (Second Ed.). New York: Wiley. [720]

[4]The actual sample size is the number of response configurations $\mathbf{y}_i$ such that $0 < y_{i+} < T$. These response configurations contain information on the structural parameters and contribute to $\ell(\boldsymbol{\theta})$; see equation (12).

ANDERSEN, E. B. (1970): "Asymptotic Properties of Conditional Maximum-Likelihood Estimators," *Journal of Royal Statistical Society, Ser. B*, 32, 283–301. [722]

ARELLANO, M., AND B. HONORÉ (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, Vol. V, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland. [719]

BARNDORFF-NIELSEN, O. (1978): *Information and Exponential Families in Statistical Theory*. New York: Wiley. [729]

BARTOLUCCI, F., AND V. NIGRO (2010): "Supplement to 'A Dynamic Model for Binary Panel Data With Unobserved Heterogeneity Admitting a $\sqrt{n}$-Consistent Conditional Estimator'," *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/7531_data.pdf; http://www.econometricsociety.org/ecta/Supmat/7531_data and programs.zip. [720]

CARRO, J. M. (2007): "Estimating Dynamic Panel Data Discrete Choice Models With Fixed Effects," *Journal of Econometrics*, 140, 503–528. [722,731]

CHAMBERLAIN, G. (1985): "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer. Cambridge: Cambridge University Press. [722]

——— (1993): "Feedback in Panel Data Models," Unpublished Manuscript, Department of Economics, Harvard University. [730]

COX, D. R. (1958): "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society, Ser. B*, 20, 215–242. [722]

——— (1972): "The Analysis of Multivariate Binary Data," *Applied Statistics*, 21, 113–120. [719, 723]

DIGGLE, P. J., P. J. HEAGERTY, K.-Y. LIANG, AND S. L. ZEGER (2002): *Analysis of Longitudinal Data* (Second Ed.). New York: Oxford University Press. [725]

HAHN, J., AND W. NEWEY (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319. [722]

HECKMAN, J. J. (1981a): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data With Econometric Applications*, ed. by D. McFadden and C. F. Manski. Cambridge, MA: MIT Press. [719]

——— (1981b): "Heterogeneity and State Dependence," in *Structural Analysis of Discrete Data With Econometric Applications*, ed. by D. McFadden and C. F. Manski. Cambridge, MA: MIT Press. [719]

——— (1981c): "The Incidental Parameter Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data With Econometric Applications*, ed. by D. McFadden and C. F. Manski. Cambridge, MA: MIT Press. [725]

HONORÉ, B. E., AND E. KYRIAZIDOU (2000): "Panel Data Discrete Choice Models With Lagged Dependent Variables," *Econometrica*, 68, 839–874. [720,722,723,728,730,731]

HONORÉ, B. E., AND E. TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629. [722,730]

HSIAO, C. (2005): *Analysis of Panel Data* (Second Ed.). New York: Cambridge University Press. [719]

HYSLOP, D. R. (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67, 1255–1294. [725]

MAGNAC, T. (2004): "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica*, 72, 1859–1876. [722]

MOLENBERGHS, G., AND G. VERBEKE (2004): "Meaningful Statistical Model Formulations for Repeated Measures," *Statistica Sinica*, 14, 989–1020. [725]

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. L. McFadden. Amsterdam: North-Holland. [730]

RASCH, G. (1961): "On General Laws and the Meaning of Measurement in Psychology," in *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, Berkeley, CA: University of California Press, 321–333. [720,722]

WOOLDRIDGE, J. M. (2001): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press. [724]

Dipartimento di Economia, Finanza e Statistica, Università di Perugia, Via A. Pascoli 20, 06123 Perugia, Italy; bart@stat.unipg.it

and

Dipartimento di Studi Economico-Finanziari e Metodi Quantitativi, Università di Roma "Tor Vergata," Via Columbia 2, 00133 Roma, Italy; Valentina.Nigro@uniroma2.it.

Manuscript received October, 2007; final revision received September, 2009.

Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lecr20

Testing for State Dependence in Binary Panel Data with Individual Covariates by a Modified Quadratic Exponential Model

Francesco Bartolucci[a], Valentina Nigro[b] & Claudia Pigini[a]
[a] University of Perugia (IT)
[b] Bank of Italy (IT)
Accepted author version posted online: 29 Jun 2015.

Testing for State Dependence in Binary Panel Data with Individual Covariates by a Modified Quadratic Exponential Model

Francesco Bartolucci[1], Valentina Nigro[2], and Claudia Pigini[1]

[1]University of Perugia (IT)

[2]Bank of Italy (IT)

Abstract

We propose a test for state dependence in binary panel data with individual covariates. For this aim, we rely on a quadratic exponential model in which the association between the response variables is accounted for in a different way with respect to more standard formulations. The level of association is measured by a single parameter that may be estimated by a Conditional Maximum Likelihood (CML) approach. Under the dynamic logit model, the conditional estimator of this parameter converges to zero when the hypothesis of absence of state dependence is true. This allows us to implement a *t*-test for this hypothesis which may be very simply performed and attains the nominal significance level under several structures of the individual covariates. Through an extensive simulation study, we find that our test has good finite sample properties and it is more robust to the presence of (autocorrelated) covariates in the model specification in comparison with other existing testing procedures for state dependence. The proposed approach is illustrated by two empirical applications: the first is based on data coming from the Panel Study of Income Dynamics and concerns employment and fertility; the second is based on the Health and Retirement Study and concerns the self reported health status.

KEYWORDS: Conditional inference; Dynamic logit model; Quadratic exponential model; t-test.

1. INTRODUCTION

In the analysis of panel data, a question of main interest is whether the choice (or the condition) of an individual in the current period may influence his/her future choice (or condition), either directly (via the so–called "true state dependence") or through the presence of unobserved time-invariant heterogeneity; see Feller (1943) and Heckman (1981). Different policy consequences may derive from disentangling the individual unobserved heterogeneity from the true state dependence, where idiosyncratic shocks may last for a long time. In the case of binary panel data, a very relevant model in which these effects are disentangled is the dynamic logit model (for a review, see Hsiao, 2005, ch. 7). This model includes individual-specific intercepts and, further to time-constant and/or time-varying individual covariates, the lagged response variable. In particular, the regression coefficient for the lagged response is a measure of the true state dependence.

The problem of modeling and testing for state dependence arises in many microeconomic applications dealing with labor market participation (Heckman and Borjas, 1980; Hyslop, 1999), transitions in and out of poverty (Cappellari and Jenkins, 2004; Biewen, 2009), and self-assessed health condition (Halliday, 2008; Heiss, 2011); lately, the problem of modeling state dependence has been raised in applications concerning migrants remittances (Bettin and Lucchetti, 2012), households financial distress (Brown et al., 2012; Giarda, 2013), and firms' access to credit (Pigini et al., 2014).

A drawback of the dynamic logit model, with respect to the static logit model that does not include the lagged response among the covariates, is that simple sufficient statistics do not exist for the individual-specific intercepts. Therefore, conditional likelihood inference becomes more complex and may be performed only under certain conditions on the distribution of the covariates (Chamberlain, 1993; Honoré and Kyriazidou, 2000). We recall that the main advantage of this approach, as other fixed-effects approaches, is that it does not require to formulate any assumption on the distribution of the individual intercepts and on the correlation between these effects and the covariates; assumptions of this type are instead required within the random-effects approach.

In order to test for (true) state dependence, Halliday (2007) found an identification condition for the case of two time periods (further to an initial observation). Nevertheless this result cannot be easily applied with longer panel settings and it does not explicitly allow for individual covariates. In particular, when covariates are present, they can be accounted for only by splitting the overall sample in strata corresponding to different configurations of these covariates, but this makes the procedure more complex and its results depending on arbitrary choices. At least to our knowledge, no other approaches having a complexity comparable to that of Halliday (2007) exist in the literature for testing for state dependence.

In this paper, we propose a test for state dependence based on a modified version of the quadratic exponential model of Bartolucci and Nigro (2010), which relies on a different formulation of the structure regarding the conditional association between the response variables given the individual-specific intercepts for the unobserved heterogeneity and the covariates. We show that the proposed model may be still represented as a latent index model where the errors are logistically distributed and the systematic part is formulated in terms of future expectations. This model may be estimated in a simple way by a Conditional Maximum Likelihood (CML) approach based on the same sufficient statistics as for the original quadratic exponential model. Moreover, we show that, when data are generated from the dynamic logit model, the estimator of the parameter measuring the association between the response variables converges to zero in absence of state dependence even in the presence of covariates. It is then natural to test for state dependence on the basis of the proposed quadratic exponential model by a t-statistic.

The test we propose is directly comparable with both the one based on Bartolucci and Nigro (2010) model and the one of Halliday (2007) in terms of simplicity of implementation. Differently from the first, our test is more powerful as it uses a larger set of information. With respect to the second test, it is unbiased and has more power in the presence of individual covariates and for panel settings of length greater than two. In addition, we show that, in the special case of two time periods and no individual covariates, the procedure we propose here and that proposed by Halliday (2007) employ the same information in the data to test for state dependence. These properties are confirmed by a deep simulation study. This study also shows that our test has a certain degree of robustness with respect to distributional misspecification of the error terms. We also extend the proposed test to accommodate for predetermined covariates in the model specification by a simple two-step procedure based on a weighting function, modeling the process of these covariates.

With the aim of illustrating the proposed test, we consider two applications. The first is about the relationship between employment and fertility and is based on a data about a sample of women which derives from the Panel Study of Income Dynamics (PSID), public use dataset.[1] The purpose is to verify, by the proposed test, well-known results on state dependence in employment and fertility (Hyslop, 1999; Carrasco, 2001; Bartolucci and Farcomeni, 2009). In the second application, we test for state dependence in the Self Reported Health Status (SRHS) using data from the Health and Retirement Study (HRS). In particular this second application is based on RAND HRS Data, Version N.[2] Using these data, Heiss (2011) found that past SRHS has a positive predictive power for the current health condition. With this example, we provide further evidence on the state dependence effect in perceived health, in addition to the cases recently analyzed by Halliday (2008) with PSID data, and by Carro and Traferri (2012) with data coming from the British Household Panel Survey.

[1]Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2005); see http://psidonline.isr.umich.edu.

[2]Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (September 2014); see http://www.rand.org/about.html.

The paper is organized as follows. In Section 2 we describe the dynamic logit model and the alternative quadratic exponential model of Bartolucci and Nigro (2010); for the purpose of our comparison, in the same section we also illustrate Halliday (2007)'s testing approach. In Section 3 we introduce the proposed t-test for state dependence based on a new formulation of the quadratic exponential model. The empirical size and the power of this test are studied by simulation in Section 4. Finally, in Section 5 we provide two empirical illustrations based on the PSID and HRS datasets. In the last section we draw the main conclusions.

We make available to the reader our `R` implementation of all the algorithms illustrated in this paper, and in particular of the algorithm to perform the proposed test for state dependence.[3] Moreover, we make available the `Stata` module `CQUAD` ("CQUAD: Stata module to perform conditional maximum likelihood estimation of quadratic exponential models").[4]

2. PRELIMINARIES

For a panel of n subjects observed at T time occasions, let y_{it} denote the binary response variable for subject i at occasion t and let $\boldsymbol{x}_{it}$ denote the corresponding vector of individual covariates. Also let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ denote the vector of all outcomes for subject i

[3] `R` package downloadable from http://cran.r-project.org/web/packages/cquad/index.html

[4] The `Stata` module `CQUAD` is downloadable from http://ideas.repec.org/c/boc/bocode/s457891.html

and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1} \cdots \boldsymbol{x}_{iT})$ denote the matrix of all covariates for this subject that are initially assumed to be strictly exogenous.

In the following, we briefly review the dynamic logit model for these data and then the quadratic exponential model as an alternative model that includes a state dependence parameter (Bartolucci and Nigro, 2010). We also review the test for state dependence proposed by Halliday (2007).

2.1. Dynamic Logit Model

The dynamic logit model assumes that, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, the binary response y_{it} has conditional distribution

$$p(y_{it} \mid \alpha_i, \boldsymbol{X}_i, y_{i0}, \ldots, y_{i,t-1}) = p(y_{it} \mid \alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}), \tag{1}$$

with probability function

$$p(y_{it} \mid \alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}) = \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]}{1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)}, \tag{2}$$

where $\boldsymbol{\beta}$ and γ are the parameters of interest and the individual-specific intercepts α_i are often considered as nuisance parameters; moreover, the initial observation y_{i0} is considered as given. Therefore, the joint probability of $\boldsymbol{y}_i$ given α_i, $\boldsymbol{X}_i$, and y_{i0} has expression

$$p(\boldsymbol{y}_i \mid \alpha_i, \boldsymbol{X}_i, y_{i0}) = \prod_t p(y_{it} \mid \alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}) = \frac{\exp(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i*}\gamma)}{\prod_t[1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]}, \tag{3}$$

where $y_{i+} = \sum_t y_{it}$ and $y_{i*} = \sum_t y_{i,t-1}y_{it}$, with the product $\prod_t$ and the sum $\sum_t$ ranging over $t = 1, \ldots, T$.

It is important to stress that γ measures the effect of the true state dependence and then the hypothesis on which we focus is $H_0 : \gamma = 0$, meaning absence of this form of dependence. The parameter γ may be identified and consistently estimated if the α_i parameters are properly taken into account. In particular, Chamberlain (1985) showed that a conditional approach, in the case of no covariates, may identify the state dependence parameter by using suitable sufficient statistics for the α_i parameters.

Honoré and Kyriazidou (2000) extended the conditional estimator of Chamberlain (1985) to allow for exogenous covariates in the model. Under particular conditions on the support of the covariates, they showed that, for $T = 3$, $\boldsymbol{y}_i$ is conditionally independent of α_i given the initial and final observations of the response variable and that $y_{i1} + y_{i2} = 1$. Their estimator has the advantage, as in any fixed-effects estimator, to let α_i be freely correlated with the covariates in $\boldsymbol{X}_i$ and to avoid any parametric assumption on the distribution of these effects. The approach relies on a weight that is attached to each observation; this weight depends on the covariates through a kernel function which reduces the rate of convergence of the estimator, which is slower than $\sqrt{n}$. Also due to this condition, the sample size substantially shrinks, lowering the overall efficiency of the estimator. Moreover, this approach does not allow for time-dummies or trend variables and may be applied only with $T > 2$, beyond the initial observation.

A different fixed-effects approach is based on bias corrected estimators; see Hahn and Newey (2004), Carro (2007), Fernandez-Val (2009), Hahn and Kuersteiner (2011), and Bartolucci et al. (2014b). These estimators are only consistent as $T \to \infty$ but have a reduced order of bias and they remain asymptotically efficient. For this reason, these estimators perform well also in quite short panels.

2.2. Quadratic Exponential Model

The quadratic exponential model directly defines the joint probability of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$ and y_{i0}, and also given an individual-specific effect here denoted by δ_i, as follows:

$$p(\boldsymbol{y}_i \mid \delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi} + y_{i*}\psi)}{\sum_{\boldsymbol{z}} \exp(z_+\delta_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + z_{i*}\psi)}, \quad (4)$$

where the sum $\sum_{\boldsymbol{z}}$ ranges over all the possible binary response vectors $\boldsymbol{z} = (z_1, \ldots, z_T)'$, $z_+ = \sum_t z_t$, and $z_{i*} = y_{i0}z_1 + \sum_{t>1} z_{t-1}z_t$. For instance, with $T = 2$, the possible vectors $\boldsymbol{z}$ are $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$, corresponding to $z_+ = 0, 1, 1, 2$ and $z_{i*} = 0, 0, y_{i0}, 1 + y_{i0}$, respectively. We refer to this model as QE1. Here we use a different notation from that used for the dynamic logit model, where the vector of regression coefficients is denoted by $\boldsymbol{\phi}$ and the state dependence parameter by ψ. These parameters are collected in the vector $\boldsymbol{\theta} = (\boldsymbol{\phi}', \psi)'$.

Model QE1 is a special case of that proposed by Bartolucci and Nigro (2010) with the parameters $\boldsymbol{\phi}$ assumed to be equal for all time occasions.[5] The same results of their paper

[5]Bartolucci and Nigro (2010) used a different parametrization for $t = T$ in order to approximate the probability in the last time period, where future covariates cannot be observed.

apply straight to model (4); therefore, the conditional probability of y_{it} may be represented as a latent index logistic model where

$$p(y_{it} \mid \delta_i, \boldsymbol{X}_i, y_{i0}, \ldots, y_{i,t-1}) = \frac{\exp\{y_{it}[\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + y_{i,t-1}\psi + e_t(\delta_i, \boldsymbol{X}_i)]\}}{1 + \exp[\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + y_{i,t-1}\psi + e_t(\delta_i, \boldsymbol{X}_i)]},$$

where, for $t = 1, \ldots, T-1$ we have

$$e_t(\delta_i, \boldsymbol{X}_i) = \log \frac{1 + \exp[\delta_i + \boldsymbol{x}'_{i,t+1}\boldsymbol{\phi} + e_{t+1}(\delta_i, \boldsymbol{X}_i) + \psi]}{1 + \exp[\delta_i + \boldsymbol{x}'_{i,t+1}\boldsymbol{\phi} + e_{t+1}(\delta_i, \boldsymbol{X}_i)]} = \log \frac{p(y_{i,t+1} = 0 \mid \delta_i, \boldsymbol{X}_i, y_{it} = 0)}{p(y_{i,t+1} = 0 \mid \delta_i, \boldsymbol{X}_i, y_{it} = 1)},$$

with

$$e_T(\delta_i, \boldsymbol{X}_i) = 1.$$

Furthermore, the quadratic exponential model shares with the dynamic logit the same interpretation of the state dependence parameter as log-odds ratio between any pair of response variables $(y_{i,t-1}, y_{it})$; moreover, y_{it} is conditionally independent of any other response variable given $y_{i,t-1}$ and $y_{i,t+1}$. Actually, when $\psi = 0$ this model coincides with the static logit model, and this is an important point for the approach here proposed.

The main advantage of model QE1 defined above is the availability of simple sufficient statistics for the unobserved heterogeneity parameters. In particular, the sufficient statistic for each parameter δ_i is y_{i+}. Therefore, a $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\phi}}', \hat{\psi})'$ may be obtained by maximizing the conditional likelihood through a Newton-Raphson algorithm. Moreover, differently from Honoré and Kyriazidou (2000), it allows for time-dummies and can be used even with $T = 2$.

Bartolucci and Nigro (2012) also showed that, up to a correction term, a quadratic exponential model of the type above may sharply approximate the dynamic logit model. On the basis of this result they derived a Pseudo CML estimator (PCML) which is very competitive in terms of efficiency compared with the other estimators proposed in the econometric literature.

2.3. Available Test for State Dependence

Halliday (2007) proposed a test for state dependence allowing for the presence of aggregate time variables in a dynamic logit model based on assumptions (1) and (2). The proposed approach, which follows the lines of the conditional approach of Chamberlain (1985), is based on the construction of conditional probability inequalities that depend only on the sign of the state dependence parameter γ.

In the case of $T = 2$, Halliday (2007) considered the events $A_{i1} = \{y_{i0} = 1, y_{i1} = 1, y_{i2} = 0\}$ and $B_{i1} = \{y_{i0} = 0, y_{i1} = 1, y_{i2} = 0\}$ and he proved that

$$p(A_{i1} \mid \boldsymbol{X}_i, y_{i0} = 1, y_{i+} = 1) \geq p(B_{i1} \mid \boldsymbol{X}_i, y_{i0} = 0, y_{i+} = 1) \qquad \text{for } \gamma \geq 0$$

and

$$p(A_{i1} \mid \boldsymbol{X}_i, y_{i0} = 1, y_{i+} = 1) \leq p(B_{i1} \mid \boldsymbol{X}_i, y_{i0} = 0, y_{i+} = 1) \qquad \text{for } \gamma \leq 0,$$

under assumption (1). When $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$ are constant across subjects, and therefore there are only time-dummies common to all sample units, it is possible to consistently estimate

$p_A = p(A_{i1} \mid \boldsymbol{X}_i, y_{i0} = 1, y_{i+} = 1)$ and $p_B = p(B_{i1} \mid \boldsymbol{X}_i, y_{i0} = 0, y_{i+} = 1)$ as follows

$$\hat{p}_A = \frac{\sum_i 1\{A_{i1} \mid \boldsymbol{X}_i\}}{\sum_i 1\{y_{i+} = 1, y_{i0} = 1 \mid \boldsymbol{X}_i\}} = \frac{n_{110}}{m_1},$$
$$\hat{p}_B = \frac{\sum_i 1\{B_{i1} \mid \boldsymbol{X}_i\}}{\sum_i 1\{y_{i+} = 1, y_{i0} = 0 \mid \boldsymbol{X}_i\}} = \frac{n_{010}}{m_0},$$

where $1\{\cdot\}$ is the indicator function, $n_{y_0 y_1 y_2}$ is the frequency of sample units with response configuration (y_0, y_1, y_2), and $m_{y_0} = n_{y_0 01} + n_{y_0 10}$. The test statistic for $H_0 : \gamma = 0$ is then defined as

$$S = \sqrt{n}\frac{\hat{p}_A - \hat{p}_B}{\hat{\sigma}(\hat{p}_A - \hat{p}_B)}, \tag{5}$$

where $\hat{\sigma}(\hat{p}_A - \hat{p}_B)$ is the estimated standard deviation of the numerator.

As the sample size grows to infinity, the distribution of the above test statistic converges to a standard normal distribution only under H_0; otherwise it diverges to $+\infty$ or to $-\infty$, as n grows to infinity, according to whether $\gamma > 0$ or $\gamma < 0$. It is worth noting that this statistic exploits all the possible configurations of the response variable, conditionally on $y_{i+} = 1$; in fact, after some simple algebra, the numerator of (5) may be written as

$$\hat{p}_A - \hat{p}_B = \frac{n_{001} n_{110} - n_{101} n_{010}}{m_1 m_0}. \tag{6}$$

The method of Halliday (2007) identifies the sign of the state dependence parameter without estimating γ and avoiding distributional assumptions on the unobserved heterogeneity parameters. Nevertheless, this result cannot be easily generalized to $T > 2$. A possible solution may be using a multiple testing technique (Hochberg and Tamhane, 1987), where tests for all possible triples $(y_{i,t-1}, y_{it}, y_{i,t+1})$, $t = 1, \ldots, T-1$, are combined together. Furthermore, to take into account individual covariates that vary across individuals and/or time occasions, the test must be performed for different configurations of the covariates. Consequently, the results may depend on how the covariate configurations are grouped and may give a final ambiguous solution.

3. A MODIFIED VERSION OF THE QUADRATIC EXPONENTIAL MODEL FOR TESTING STATE DEPENDENCE

In the following, we first illustrate a modified version of the quadratic exponential model QE1 outlined in Section 2.2 and then we discuss how to estimate its parameters. On the basis of these estimates we introduce our test for state dependence.

3.1. Modified Quadratic Exponential Model

We introduce a different version of the quadratic exponential model (4), which is defined as

$$\tilde{p}(\boldsymbol{y}_i \mid \delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{y}_{i*}\psi)}{\sum_z \exp(z_+\delta_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{z}_{i*}\psi)}, \tag{7}$$

where

$$\tilde{y}_{i*} = \sum_t 1\{y_{it} = y_{i,t-1}\},$$

$$\tilde{z}_{i*} = 1\{z_1 = y_{i0}\} + \sum_{t>1} 1\{z_t = z_{t-1}\},$$

and we use $\tilde{p}(\cdot\,|\,\cdot)$ instead of $p(\cdot\,|\,\cdot)$ for the probability function in order to avoid confusion, given that the two models use the same parameters; in particular, the parameters of interest are still collected in the vector $\boldsymbol{\theta}$.

The model based on expression (7) is referred to as QE2. The main difference between models QE1 and QE2 is in how the association between the response variables is formulated. In the latter, the structure is based on the statistic $\tilde{y}_{i*}$ that, differently from y_{i*}, is equal to the number of consecutive pairs of outcomes which are equal each other, regardless if they are 0 or 1. As already mentioned, this allows us to use a larger set of information with respect to the initial model QE1 in testing for state dependence; this issue will be discussed in more detail in the following.

Regarding the interpretation of model QE2, it is useful to consider how expression (7) becomes after recursive marginalizations of the response variables in backward order. In particular, for $t = 1, \ldots, T-1$, we have that

$$\tilde{p}(y_{i1}, \ldots, y_{it}\,|\,\boldsymbol{X}_i, y_{i0}) = \frac{\exp(\sum_{h=1}^{t} y_{ih}\delta_i + \sum_{h=1}^{t} y_{ih}\boldsymbol{x}'_{ih}\boldsymbol{\phi} + \sum_{h=1}^{t} 1\{y_{ih} = y_{i,h-1}\}\psi)\tilde{g}_t(y_{it}, \delta_i, \boldsymbol{X}_i)}{\sum_{\boldsymbol{z}} \exp(z_+\delta_i + \sum_{h=1}^{T} z_h\boldsymbol{x}'_{ih}\boldsymbol{\phi} + \tilde{z}_{i*}\psi)},$$

where

$$\tilde{g}_t(y_{it}, \delta_i, \boldsymbol{X}_i) = \tilde{g}_{t+1}(0, \delta_i, \boldsymbol{X}_i)\exp[(1 - y_{it})\psi] + \tilde{g}_{t+1}(1, \delta_i, \boldsymbol{X}_i)\exp(\delta_i + \boldsymbol{x}'_{i,t+1}\boldsymbol{\phi} + y_{it}\psi),$$

with $\tilde{g}_T(y_{iT}, \delta_i, \boldsymbol{X}_i) = 1$. Consequently we have that

$$\log \frac{\tilde{p}(y_{i1}, \ldots, y_{i,t-1}, y_{it} = 1 \,|\, \boldsymbol{X}_i, y_{i0})}{\tilde{p}(y_{i1}, \ldots, y_{i,t-1}, y_{it} = 0 \,|\, \boldsymbol{X}_i, y_{i0})} = \delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + (2y_{i,t-1} - 1)\psi + \tilde{e}_t(\delta_i, \boldsymbol{X}_i),$$

where

$$\begin{aligned}
\tilde{e}_t(\delta_i, \boldsymbol{X}_i) &= \log \frac{\tilde{g}_t(1, \delta_i, \boldsymbol{X}_i)}{\tilde{g}_t(0, \delta_i, \boldsymbol{X}_i)} \\
&= \log \frac{\tilde{g}_{t+1}(0, \delta_i, \boldsymbol{X}_i) + \tilde{g}_{t+1}(1, \delta_i, \boldsymbol{X}_i)\exp(\delta_i + \boldsymbol{x}'_{i,t+1}\boldsymbol{\phi} + \psi)}{\tilde{g}_{t+1}(0, \delta_i, \boldsymbol{X}_i)\exp(\psi) + \tilde{g}_{t+1}(1, \delta_i, \boldsymbol{X}_i)\exp(\delta_i + \boldsymbol{x}'_{i,t+1}\boldsymbol{\phi})} \\
&= \log \frac{1 + \exp[\delta_i + \boldsymbol{x}'_{i,t+1}\boldsymbol{\phi} + \psi + \tilde{e}_{t+1}(\delta_i, \boldsymbol{X}_i)]}{1 + \exp[\delta_i + \boldsymbol{x}'_{i,t+1}\boldsymbol{\phi} - \psi + \tilde{e}_{t+1}(\delta_i, \boldsymbol{X}_i)]} - \psi, \qquad (8)
\end{aligned}$$

for $t = 1, \ldots, T-1$, with $\tilde{e}_T(\delta_i, \boldsymbol{X}_i) = 0$. This implies that

$$\tilde{p}(y_{iT} \,|\, \delta_i, \boldsymbol{X}_i, y_{i,T-1}) = \frac{\exp\{y_T[\delta_i + \boldsymbol{x}'_{iT}\boldsymbol{\phi} + (2y_{i,T-1} - 1)\psi]\}}{1 + \exp[\delta_i + \boldsymbol{x}'_{iT}\boldsymbol{\phi} + (2y_{i,T-1} - 1)\psi]}. \qquad (9)$$

This expression may be seen as a reparametrization of the probability expression holding under the dynamic logit model (1); in this regard, it is useful to recall that $(2y_{i,t-1} - 1)$ is simply equal to -1 for $y_{i,t-1} = 0$ and to 1 for $y_{i,t-1} = 1$. For $t = 1, \ldots, T-1$, instead, we have

$$\tilde{p}(y_{it} \,|\, \delta_i, \boldsymbol{X}_i, y_{i,t-1}) = \frac{\exp\{y_t[\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + (2y_{i,t-1} - 1)\psi + \tilde{e}_t(\delta_i, \boldsymbol{X}_i)]\}}{1 + \exp[\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + (2y_{i,t-1} - 1)\psi + \tilde{e}_t(\delta_i, \boldsymbol{X}_i)]}. \qquad (10)$$

Regarding the last expression, first of all consider that for $\psi = 0$ definition (8) implies that $\tilde{e}_t(\delta_i, \boldsymbol{X}_i) = 0$ and then we have again a reparametrization of the dynamic logit model. Moreover, we have that

$$\tilde{e}_t(\delta_i, \boldsymbol{X}_i) = \log \frac{\tilde{p}(y_{i,t+1} = 0 \,|\, \delta_i, \boldsymbol{X}_i, y_{it} = 0)}{\tilde{p}(y_{i,t+1} = 0 \,|\, \delta_i, \boldsymbol{X}_i, y_{it} = 1)} - \psi,$$

which directly compares with the expression of $e_t(\delta_i, \boldsymbol{X}_i)$ given in Section 2.2 for model QE1. This may easily be proved by recognizing that the numerator and the denominator in (8) are equal to the inverse of $\tilde{p}(y_{i,t+1} = 0 \mid \delta_i, \boldsymbol{X}_i, y_{it} = 1)$ and $\tilde{p}(y_{i,t+1} = 0 \mid \delta_i, \boldsymbol{X}_i, y_{it} = 0)$, respectively, as defined in (9) and (10).

The above correction term depends on the data only through $\boldsymbol{x}_{i,t+1}, \dots, \boldsymbol{x}_{iT}$ and has an interpretation in terms of the probability of future choices similar to that of model QE1 and the quadratic exponential model of Bartolucci and Nigro (2010).

It may simply be proved model QE2 reproduces the same conditional independence relations of the dynamic logit model between the response variable y_{it} and $y_{i0}, \dots, y_{i,t-2}$, $y_{i,t+2}, \dots, y_{iT}$, given α_i, $\boldsymbol{X}_i$, $y_{i,t-1}$, and $y_{i,t+1}$ ($t = 2, \dots, T-1$). Finally, we have

$$\log \frac{\tilde{p}(y_{it} = 1 \mid \delta_i, \boldsymbol{X}_i, y_{i,t-1} = 1)}{\tilde{p}(y_{it} = 0 \mid \delta_i, \boldsymbol{X}_i, y_{i,t-1} = 1)} \frac{\tilde{p}(y_{it} = 0 \mid \delta_i, \boldsymbol{X}_i, y_{i,t-1} = 0)}{\tilde{p}(y_{it} = 1 \mid \delta_i, \boldsymbol{X}_i, y_{i,t-1} = 0)} = 2\psi, \qquad t = 1, \dots, T,$$

meaning that the log-odds ratio between every consecutive pair of response variables has the same sign of ψ and it is equal to 0 if there is no state dependence.

3.2. Model Estimation

As for model QE1, the sums of the response variables at the individual level, y_{i+}, $i = 1, \dots, n$, are sufficient statistics for the individual-specific intercepts δ_i. Conditioning on the sum of the response variables, we obtain for model QE2 the following conditional probability function:

$$\tilde{p}(\boldsymbol{y}_i \mid \boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{y}_{i*}\psi)}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{z}_{i*}\psi)}. \qquad (11)$$

On the basis of expression (11), we obtain the conditional log-likelihood

$$\tilde{\ell}(\boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\}\tilde{\ell}_i(\boldsymbol{\theta}), \qquad (12)$$

where

$$\begin{aligned}\tilde{\ell}_i(\boldsymbol{\theta}) &= \log \tilde{p}(\boldsymbol{y}_i \mid \boldsymbol{X}_i, y_{i0}, y_{i+})\\ &= \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{y}_{i*}\psi - \log \sum_{\boldsymbol{z}:z_+=y_{i+}} \exp\left(\sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{z}_{i*}\psi\right) \qquad (13)\end{aligned}$$

is the individual contribution to the conditional log-likelihood. Note that the response configurations with y_{i+} equal to 0 or T do not contribute to this likelihood and then they are not considered in (12).

Function $\tilde{\ell}(\boldsymbol{\theta})$ may be maximized by a Newton-Raphson algorithm in a similar way as for model QE1, using the score vector and the information matrix reported below; see also Bartolucci and Nigro (2010). In this regard, it is convenient to write

$$\tilde{\ell}_i(\boldsymbol{\theta}) = \tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)'\boldsymbol{\theta} - \log \sum_{\boldsymbol{z}:z_+=y_{i+}} \exp[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z})'\boldsymbol{\theta}],$$

with

$$\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) = \left(\sum_t y_{it}\boldsymbol{x}'_{it}, \tilde{y}_{i*}\right)',$$

so that, using the standard theory about the regular exponential family, we have the following expressions for the score for $\tilde{\ell}(\boldsymbol{\theta})$:

$$\tilde{\boldsymbol{s}}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\,\tilde{\ell}(\boldsymbol{\theta}) = \sum_i \tilde{\boldsymbol{s}}_i(\boldsymbol{\theta}),$$
$$\tilde{\boldsymbol{s}}_i(\boldsymbol{\theta}) = \tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) - \widetilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}].$$

Regarding the observed information matrix we have

$$\widetilde{\boldsymbol{J}}(\boldsymbol{\theta}) = -\nabla_{\boldsymbol{\theta\theta}}\,\tilde{\ell}(\boldsymbol{\theta}) = \sum_i \tilde{V}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}]. \tag{14}$$

In these expressions, $\widetilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}]$ denotes the conditional expected value of $\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)$ given $\boldsymbol{X}_i$ and y_{i+} under model QE2, whereas the corresponding conditional variance is denoted by $\tilde{V}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}]$. These are given by

$$\widetilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}] = \sum_{\boldsymbol{z}: z_+ = y_{i+}} \tilde{p}(\boldsymbol{z} \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+})\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z})$$
$$\widetilde{\mathrm{V}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}] = \sum_{\boldsymbol{z}: z_+ = y_{i+}} \tilde{p}(\boldsymbol{z} \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+})\tilde{\boldsymbol{d}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z})\tilde{\boldsymbol{d}}(\boldsymbol{X}, y_{i0}, \boldsymbol{z})',$$

with $\tilde{\boldsymbol{d}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z}) = \tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z}) - \widetilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}]$.

Note that $\widetilde{\boldsymbol{J}}(\boldsymbol{\theta})$ defined above is always non-negative definite since it corresponds to the sum of a series of variance-covariance matrices and therefore $\tilde{\ell}(\boldsymbol{\theta})$ is always concave. Moreover, a necessary condition for the information matrix $\widetilde{\boldsymbol{J}}(\boldsymbol{\theta})$ to be non-singular, and then for

$\tilde{\ell}(\boldsymbol{\theta})$ to be strictly concave, is that time-constant covariates are ruled out, as happens in any fixed-effect approach. The CML estimator of $\boldsymbol{\theta}$ based on the maximization of (12) is denoted by $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\phi}}', \tilde{\psi})'$.

3.3. Testing for State Dependence

Once the parameters of the proposed quadratic exponential model are estimated, it is straightforward to construct a standard t-statistic for testing $H_0 : \psi = 0$, as follows:

$$W = \frac{\tilde{\psi}}{\mathrm{se}(\tilde{\psi})}, \tag{15}$$

where se$(\cdot)$ is the standard error derived using the sandwich estimator of White (1982). In particular, from the log-likelihood equation defined in (13), the variance-covariance matrix of $\tilde{\boldsymbol{\theta}}$ is estimated as

$$\widetilde{\boldsymbol{V}}(\tilde{\boldsymbol{\theta}}) = \widetilde{\boldsymbol{J}}(\tilde{\boldsymbol{\theta}})^{-1}\widetilde{\boldsymbol{H}}(\tilde{\boldsymbol{\theta}})[\widetilde{\boldsymbol{J}}(\tilde{\boldsymbol{\theta}})^{-1}]',$$

where

$$\widetilde{\boldsymbol{H}}(\boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\}\tilde{\boldsymbol{s}}_i(\boldsymbol{\theta})\tilde{\boldsymbol{s}}_i(\boldsymbol{\theta})'$$

and $\widetilde{\boldsymbol{J}}(\boldsymbol{\theta})$ is the information matrix defined in (14). Once the matrix $\widetilde{\boldsymbol{V}}(\tilde{\boldsymbol{\theta}})$ has been computed as above, the standard error for $\tilde{\psi}$ may be obtained in the usual way from the main diagonal of this matrix.

It is worth noting that, under H_0, the dynamic logit model corresponds to the proposed quadratic exponential model QE2 when $\psi = 0$. In fact, when $\gamma = 0$, expression (3) that holds under the dynamic logit model simplifies to

$$p(\boldsymbol{y}_i \,|\, \alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\sum_z \exp(z_+\alpha_i + \sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\beta})},$$

which coincides with that in (7) holding under the model QE2, with $\psi = 0$, $\boldsymbol{\phi} = \boldsymbol{\beta}$, and $\delta_i = \alpha_i$; the same obviously happens for model QE1. This implies the following main result.

Proposition 1. *Under the dynamic logit model with strictly exogenous covariates based on assumptions* (1) *and* (2), *if the null hypothesis* $H_0 : \gamma = 0$ *holds, the test statistic* W *defined in* (15) *has asymptotic standard normal distribution as* $n \to \infty$.

Moreover, if data are generated from the dynamic logit model but $\gamma \neq 0$, then the value of W is expected to diverge to $+\infty$ or $-\infty$ according to whether the true value of γ is positive or negative. This is because, as we show at the end of Section 3.1, the sign of ψ is the same of the log-odds ratio between pairs of consecutive response variables and the latter is equal to γ under the dynamic logit model. Therefore, within the proposed approach we reject H_0 against the unidirectional alternative $H_1 : \gamma > 0$, at the significance level α, if the observed value of W is greater than z_α, where z_α is the 100α-th upper percentile of the standard normal distribution. Similarly, we reject H_0 against $H_1 : \gamma < 0$ if the observed value of W is smaller than $-z_\alpha$ and we reject H_0 against the bidirectional alternative $H_1 : \gamma \neq 0$ if the observed value of $|W|$ is greater than $z_{\alpha/2}$.

A relevant issue is if the same t-test as above may be based on the initial quadratic exponential model QE1, using a statistic of type $\hat{\psi}/se(\hat{\psi})$, as also this model is equal to the static logit model when $\psi = 0$. Our conjecture is that this test is less powerful than the test based on the test statistic W defined above because the latter is based on a version of the quadratic exponential model, the estimator of which better exploits the information about the association between the response variables.

In order to illustrate the above point, we consider the simple case in which there are two time occasions, no covariates, and no time-dummies. In this case it is possible to prove that the CML estimator of ψ under model QE1 is equal to

$$\hat{\psi} = \log \frac{n_{110}}{n_{101}}$$

in terms of sample frequencies, whereas, under model QE2, the estimator of this parameter has the explicit expression

$$\tilde{\psi} = \log \frac{n_{001} + n_{110}}{n_{010} + n_{101}}. \tag{16}$$

In fact, after some simple algebra we have

$$\tilde{\ell}(\psi) = \sum_i 1\{y_{i+} = 1\} \Big\{ (1\{y_{i1} = y_{i0}\} + 1\{y_{i2} = y_{i1}\})\psi - \log \sum_{z:z_+ = y_{i+}} \exp[(1\{z_1 = y_{i0}\} + 1\{z_2 = z_1\})\psi] \Big\},$$

that, in terms of sample frequencies, may be expressed as

$$\tilde{\ell}(\psi) = (n_{001} + n_{110})\psi - m_0 \log k(0) - m_1 \log k(1),$$

where $k(y_{i0}) = \exp[(1-y_{i0})\psi] + \exp(y_{i0}\psi) = 1 + \exp(\psi)$ and we recall that $m_{y_0} = n_{y_0 01} + n_{y_0 10}$. Consequently, the score function is

$$\tilde{s}(\psi) = \sum_i 1\{y_{i+} = 1\} \left[1\{y_{i1} = y_{i0}\} - \frac{1\{y_{i1} = y_{i0}\} \exp(1\{y_{i1} = y_{i0}\}\psi)}{k(y_{i0})} \right]$$

$$= n_{001} + n_{110} - (m_0 + m_1)\frac{\exp(\psi)}{1 + \exp(\psi)}.$$

Solving $\tilde{s}(\psi) = 0$ we obtain

$$\frac{\exp(\psi)}{1 + \exp(\psi)} = \frac{n_{001} + n_{110}}{m_0 + m_1},$$

and

$$\text{logit}\, \frac{\exp(\psi)}{1 + \exp(\psi)} = \tilde{\psi} = \log \frac{n_{001} + n_{110}}{m_0 + m_1 - (n_{001} + n_{110})},$$

which reduces to (16).

Both the above estimators converge in probability to the true value of γ under the dynamic logit model if H_0 holds. However, the first estimator exploits a reduced amount of information with respect to the second, as it ignores the response configurations $n_{y_0 y_1 y_2}$ with $y_0 = 0$; something similar happens in more complex situations. Consequently the test based on estimator $\hat{\psi}$ (QE1) attains a reduced power than the one based on the estimator

$\tilde{\psi}$ (QE2) when the true value of γ is different from 0. As will be shown in Section 4.2, this different behavior is also confirmed by the simulation study.[6]

In order to better compare our test with that of Halliday (2007), we consider the case with $T = 2$ and only time dummies illustrated in Section 2.3. The conditional log-likelihood of model QE2, as defined in (12), allows us to identify two parameters, that is, ϕ corresponding to the difference between the effect of the two time-dummies and ψ for the state dependence, so that $\boldsymbol{\theta} = (\phi, \psi)'$. Moreover, after some simple algebra similar to that used above, we have that

$$\tilde{\ell}(\boldsymbol{\theta}) = (n_{001} + n_{101})\phi + (n_{001} + n_{110})\psi - m_0 \log k(0) - m_1 \log k(1),$$

where now $k(y_{i0}) = \exp[\phi + (1 - y_{i0})\psi] + \exp(y_{i0}\psi)$. Consequently, the score function is

$$\tilde{\boldsymbol{s}}(\boldsymbol{\theta}) = \begin{pmatrix} n_{001} + n_{101} - \frac{m_0 \exp(\phi+\psi)}{k(0)} - \frac{m_1 \exp(\phi)}{k(1)} \\ n_{001} + n_{110} - \frac{m_0 \exp(\phi+\psi)}{k(0)} - \frac{m_1 \exp(\psi)}{k(1)} \end{pmatrix}.$$

In order to solve the system of two equations $\tilde{\boldsymbol{s}}(\boldsymbol{\theta}) = \boldsymbol{0}$, we initially subtract the first equation from the second and, after some algebra, we obtain

$$\exp(\psi) = \frac{n_{110}}{n_{101}} \exp(\phi). \qquad (17)$$

[6]A related point is how the proposed test compares with a t-test based on one of the fixed-effects estimators for the dynamic logit model, as the PCML estimator proposed by Bartolucci and Nigro (2012). Since this estimator is based on a model similar to QE1, as an approximating model, we expect a similar difference in terms of power with respect to the proposed test for state dependence.

We then substitute this result in the first equation obtaining

$$\tilde{\phi} = \frac{1}{2} \log \frac{n_{001} n_{101}}{n_{010} n_{110}}.$$

Finally, by substituting this solution in (17), we have

$$\tilde{\psi} = \frac{1}{2} \log \frac{n_{001} n_{110}}{n_{010} n_{101}}.$$

The last result proves that our test statistic is based on the same response variable configurations of Halliday's statistic in (6) and then it exploits the same amount of information. Moreover, the two test statistics always exhibit the same sign since

$$\mathrm{sign}(\tilde{\psi}) = \mathrm{sign}[\log(n_{001} n_{110}) - \log(n_{010} n_{101})] = \mathrm{sign}(n_{001} n_{110} - n_{010} n_{101}) = \mathrm{sign}(\hat{p}_A - \hat{p}_B),$$

where $\hat{p}_A - \hat{p}_B$ in the numerator of Halliday's test statistic S defined in (5). This also confirms that our estimator $\tilde{\psi}$ identifies the sign of the state dependence parameter γ under the dynamic logit model. A consequence is that the proposed test statistic W and the test statistic S have the same asymptotic distribution with mean centered in 0 under the dynamic logit model when $H_0 : \gamma = 0$ holds; both test statistics diverge to $+\infty$ or $-\infty$ under the dynamic logit model with $\gamma \neq 0$ (the first case when the true value of γ is positive and the second when it is negative). This is in agreement with the similar performance of the tests based on the two statistics, S and W, in terms of actual size and power that we note in the simulation study when $T = 2$ and in absence of individual covariates.

3.4. Model Estimation and Testing with Predetermined Covariates

The key of the proposed approach is that, under the dynamic logit model defined in (2), the CML estimator of the state dependence parameter ψ in model QE2, as well as in model QE1, converges in probability to 0 under $H_0 : \gamma = 0$. This result holds, regardless of the distribution of the individual-specific parameters α_i, provided that the covariates are strictly exogenous, so that (1) holds. Therefore, when there are predetermined covariates, estimation of the state dependence parameter ψ may be biased and the proposed test is not ensured to attain the nominal significance level. In the following we propose a simple correction of the proposed approach to overcome this problem.

For $t = 1, \ldots, T$, denote by $f\left(\boldsymbol{x}_{it} \mid \boldsymbol{x}_{i,t-1}, y_{i,t-1}\right)$ the conditional distribution of the covariates $\boldsymbol{x}_{it}$ given the lagged covariates and the response. In this framework, the conditional distribution of $p(\boldsymbol{y}_i \mid \alpha_i, \boldsymbol{x}_{i0}, \boldsymbol{X}_i, y_{i0})$ under the dynamic logit model is different from that given in (3). In fact, we have

$$f(\boldsymbol{y}_i, \boldsymbol{X}_i \mid \alpha_i, \boldsymbol{x}_{i0}, y_{i0}) = \prod_t p(y_{it} \mid \alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}) f(\boldsymbol{x}_{it} \mid \boldsymbol{x}_{i,t-1}, y_{i,t-1}),$$

with $p(y_{it} \mid \alpha_i, \boldsymbol{x}_{it}, y_{i,t-1})$ is defined in (1). Consequently, we have

$$p(\boldsymbol{y}_i \mid \alpha_i, x_{i0}, \boldsymbol{X}_i, y_{i0}) = \frac{\prod_t p(y_{it} \mid \alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}) f(\boldsymbol{x}_{it} \mid \boldsymbol{x}_{i,t-1}, y_{i,t-1})}{\sum_z \prod_t p(z_t \mid \alpha_i, \boldsymbol{x}_{it}, z_{t-1}) f(\boldsymbol{x}_{it} \mid \boldsymbol{x}_{i,t-1}, z_{t-1})}$$

that under $\gamma = 0$ simplifies to

$$p(\boldsymbol{y}_i \mid \alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta}) \prod_t f(\boldsymbol{x}_{it} \mid \boldsymbol{x}_{i,t-1}, y_{i,t-1})}{\sum_z \exp(z_+\alpha_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\beta}) \prod_t f(\boldsymbol{x}_{it} \mid \boldsymbol{x}_{i,t-1}, z_{t-1})}.$$

It is clear that this expression is different from that holding under model QE2, see equation (7), when $\psi = 0$, $\boldsymbol{\phi} = \boldsymbol{\beta}$ and $\delta_i = \alpha_i$ and, therefore, Proposition 1 is not ensured to hold. It is also clear that, if we want this proposition to hold even under the predetermined framework defined above, model QE2 must be modified by including the component

$$\omega(\boldsymbol{x}_{i0}, \boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) = \prod_t f(\boldsymbol{x}_{it} \,|\, \boldsymbol{x}_{i,t-1}, y_{i,t-1}), \tag{18}$$

which may be seen as a weight associated to the response and covariate configuration.

The distribution of the response variables under the extended QE2 model is

$$\tilde{p}_\omega(\boldsymbol{y}_i \,|\, \delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\phi} + \tilde{y}_{i*}\psi)\omega(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)}{\sum_{\boldsymbol{z}} \exp(z_+\delta_i + \sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\phi} + \tilde{z}_{i*}\psi)\omega(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z})},$$

which reduces to expression (7) when covariates are strictly exogenous, which implies that $f(\boldsymbol{x}_{it} \,|\, \boldsymbol{x}_{i,t-1}, y_{i,t-1}) = f(\boldsymbol{x}_{it} \,|\, \boldsymbol{x}_{i,t-1})$. Finally, the conditional probability of $\boldsymbol{y}_i$ given y_{i+} is

$$\tilde{p}_\omega(\boldsymbol{y}_i \,|\, \boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\phi} + \tilde{y}_{i*}\psi)\omega(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\phi} + \tilde{z}_{i*}\psi)\omega(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z})}, \tag{19}$$

which directly extends (11).

On the basis of expression (19), estimation and testing of ψ can be carried out by a two-step approach. The first step consists of estimating the model for the predetermined covariates in a suitable way, so as to obtain $\hat{f}(\boldsymbol{x}_{it} \,|\, \boldsymbol{x}_{i,t-1}, y_{i,t-1})$ and then the estimated weighting function $\hat{\omega}(\boldsymbol{x}_{i0}, \boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)$ based on (18). The second step consists of maximizing the conditional log-likelihood $\tilde{\ell}_\omega(\boldsymbol{\theta})$, which is defined as in (12) on the basis of the individual contributes

$$\tilde{\ell}_{\omega,i}(\boldsymbol{\theta}) = \sum_t y_{it}\boldsymbol{x}_{it}\boldsymbol{\phi} + \tilde{y}_{i*}\psi + \log\hat{\omega}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)$$
$$- \log \sum_{\boldsymbol{z}:z_+=y_{i+}} \exp\left(\sum_t z_t\boldsymbol{x}_{it}\boldsymbol{\phi} + \tilde{z}_{i*}\psi\right)\hat{\omega}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z}).$$

The maximization of the $\tilde{\ell}_\omega(\boldsymbol{\theta})$ proceeds by a simple extension of the Newton-Raphson algorithm outlined in Section 3.2, so as to obtain the estimator $\hat{\boldsymbol{\theta}}_\omega = (\hat{\boldsymbol{\phi}}'_\omega, \hat{\psi}_\omega)'$. Then, the test statistic can be computed as $W_\omega = \tilde{\psi}_\omega/\text{se}(\tilde{\psi}_\omega)$, where $\text{se}(\tilde{\psi}_\omega)$ is corrected as in Murphy and Topel (1985), and under H_0 has still an asymptotic standard normal distribution even with predetermined covariates. Obviously, a limit of this approach is that the unobserved heterogeneity is not incorporated in the model for the predetermined covariates. Nevertheless, simulations will show that the bias in the test statistic is negligible in short panels.

4. SIMULATION STUDY

In order to study the finite–sample properties of the t-test for state dependence proposed in Section 3, we performed a comprehensive Monte Carlo experiment based on a simulation design similar to the one adopted by Honoré and Kyriazidou (2000).

4.1. Design

We generated samples from a dynamic logit model where the conditional mean specification includes individual-specific intercepts, one covariate, and the lag of the response variable as follows:

$$y_{it} = 1\left\{\alpha_i + x_{it}\beta + y_{i,t-1}\gamma + \varepsilon_{it} \geq 0\right\},$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, with initial condition

$$y_{i0} = 1\left\{\alpha_i + x_{i0}\beta + \varepsilon_{i0} \geq 0\right\}.$$

The error terms ε_{it} are independent, and have zero–mean logistic distribution with variance $\pi^2/3$. For $T \geq 2$, the individual intercepts α_i are obtained as $\alpha_i = \frac{1}{3}\sum_{t=0}^{2} x_{it}$, where the covariate x_{it} is generated as

$$x_{i0} \sim N\left(0, \pi^2/3\right),$$

$$x_{it} = x_{i,t-1}\rho + u_{it},$$

$$u_{it} \sim N\left(0, \left(1-\rho^2\right)\pi^2/3\right),$$

so that x_{it} and ε_{it} have the same stationary variance. In this way, the generating model admits a correlation between the covariates and the individual-specific intercepts and also it allows for an autocorrelation of the covariate for the same unit according to an AR(1) dependence structure. In particular, the covariate is autocorrelated if the parameter ρ is different from 0, whereas if ρ is equal to 0 we have a simulation design with independent covariate values at different occasions.

Based on the above generating model, we ran experiments for values of γ on a grid between -1.0 and 1.0 with step 0.1. The values of sample size we considered are $n = 500, 1000$

for $T = 2, 5$, $\beta = 0, 1$, and $\rho = 0.5$. The number of Monte Carlo replications in each scenario is $1,000$. We also considered three variations of our benchmark design. In order to investigate the properties of the t-test based on model QE2 when the distribution of ε_{it} is non-logistic, we generated ε_{it} as a standard normal random variable. In addition, we considered a static process for x_{it}, that is $\rho = 0$.

Finally, we set up a scenario where the covariate x_{it} is predetermined according to the following model:

$$x_{it} = x_{i,t-1}\rho + y_{i,t-1}\eta + v_i + u_{it}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{20}$$

where ρ and η are both equal to 0.5, u_{it} is generated as above, and v_i is a time-invariant unobserved effect. In this case, we computed the test performing a two-step estimation of model QE2 as described in Section 3.4. Since the first-step specification does not account for individual heterogeneity, we performed the simulation in the case of $v_i = 0$, $i = 1, \ldots, n$, which is the assumed data generating process. We also simulated these effects so that $v_i \sim N(0, 1)$ and $\mathrm{E}(\alpha_i v_i) = 0.5$, in order to check whether the approach is robust to the presence of unobserved heterogeneity. The first step consists of estimating, by a pooled linear regression model, the parameters ρ, η, and τ^2, with τ^2 being the variance of u_{it}, obtaining $\hat{\rho}$, $\hat{\eta}$, and $\hat{\tau}^2$, respectively. Then, according to definition (18), the weighted function is obtained as

$$\hat{\omega}(x_{i0}, \boldsymbol{x}_i, y_{i0}, \boldsymbol{y}_i) = \prod_t \varphi\left(\frac{x_{it} - (x_{i,t-1}\hat{\rho} + y_{i,t-1}\hat{\eta})}{\hat{\tau}}\right),$$

where $\varphi(\cdot)$ denotes the standard normal density function.

We performed the proposed t-test for state dependence based on model QE2 and compared its behavior with those of the t-test based on model QE1 and of Halliday's test, illustrated in Sections 2.2 and 2.3 respectively. An important feature of the latter is that it does not allow us to take into account individual covariates. A possible solution is performing the test separately for subgroups of individuals. The problem is relevant when the covariates are autocorrelated, as it is reasonable to expect in standard economic applications. A procedure that ignores the presence of these explanatory variables may confound state dependence with the persistence that comes from the correlation of $y_{i,t-1}$ with x_{it}, as both depend on $x_{i,t-1}$. Therefore, we expect Halliday's test to exhibit rather poor size properties in these circumstances.

From Halliday (2007), another issue is that it is not obvious how to test for state dependence when $T > 2$. In our simulation, we considered all the possible triples $(y_{i,t-1}, y_{it}, y_{i,t+1})$, $t = 1, \ldots, T-1$, computed Halliday's test for each of these triples and then decided when to reject $H_0 : \gamma = 0$ by a multiple testing technique (see Hochberg and Tamhane, 1987). In particular, H_0 is rejected if at least one of the p–values that are obtained from each of the $T-1$ triples of consecutive observations is smaller than the Bonferroni corrected nominal size. Such a correction ensures that the family-wise error rate is controlled for. For instance, if we test the null hypothesis for a nominal size of 0.05, the corrected nominal size is $1 - \sqrt[T-1]{0.95}$ for each single test.

4.2. Simulation Results

Figures 1 and 2 depict the power curves resulting from the simulation study. We report the results for the proposed t-test defined in (15) based on model QE2, for the corresponding test based on model QE1 defined in (4), and for the test proposed by Halliday (2007) based on statistic S defined in (5). For a better comparison with the Halliday's test, we estimated models QE2 and QE1 including the covariate x_{it} and without this covariate. For the t-tests, the curves labeled "QE2_cov" and "QE1_cov" refer to the situation in which x_{it} is included in the model specification; the curves are labeled by "QE2_nocov" and "QE1_nocov" when this covariate is ignored in the estimation. Rejection rates are displayed for a nominal size $\alpha = 0.05$ and against the bidirectional alternative hypothesis $H_1 : \gamma \neq 0$. For certain relevant values of γ, Table 1 displays the rejection rate of this bidirectional test.[7]

The first two panels of Figure 1 show that the proposed t-test has empirical size equal to the nominal level α when $\beta = 0$ and $T = 2$ and with both sample sizes, whereas a sample size of at least 1000 is needed to exhibit satisfactory power properties. The power of the test based on model QE1 is lower compared to that the proposed t-test and it increases only slightly with a sample of 1000 and $T = 2$. The test based on model QE1 performs, instead, similarly to the t-test when T is larger (see also Table 1). We recall that the main

[7]For each approach, we also considered both lower and upper tailed tests which are referred to $H_1 : \gamma < 0$ and $H_1 : \gamma > 0$, respectively. In order to limit the number of results presented, we focus the discussion of the Monte Carlo results on the bidirectional test while results for the lower and upper tailed tests are available upon request.

difference with the model QE2 here adopted for the t-test is in the way the association between the response variables is accounted for.

In these basic scenarios, Halliday's test statistic presents a behavior very similar to the proposed test. With $T = 5$ and $\beta = 0$, the rejection rate for the proposed test sensibly increases (see the third and fourth panels of Figure 1) reaching almost 100% for $|\gamma| = 0.6$ with $n = 500$ and $|\gamma| = 0.4$ with $n = 1000$. On the contrary, the generalization of Halliday's test statistic to cases with $T > 2$ leads to a remarkable power loss: with $n = 500$ and $|\gamma| = 0.5$ the rejection rate is about one half of that of the proposed test (see also Table 1).

Figure 2 provides an illustration of the simulation results with $\beta = 1$. In this case, the proposed test, "QE2_cov", and the test based on model QE1 maintain their size properties and "QE1_cov" still exhibits a power loss, compared to "QE2_cov", when T is equal to 2. In this scenario, however, Halliday's test over-rejects the null hypothesis of absence of state dependence when this hypothesis is true. Moreover, the test size bias grows with the sample size: when $T = 2$, for example, the rejection rate under H_0 rises from 23% with $n = 500$ to 44% with $n = 1000$ (see Table 1). As expected, ignoring the presence of covariates in testing for state dependence leads to mistakenly detect a significant persistence in the dependent variable.

This result is also confirmed by the rejection rates of "QE2_nocov" and "QE1_nocov" that exhibit the same behavior as Halliday's. Regarding the power, it decreases for all the three tests when $\beta = 1$ rather than $\beta = 0$. Nevertheless, our test shows a better performance in the case $\gamma < 0$. On the other hand, for $\gamma > 0$ and $T = 2$ this test has less power than Halliday's, which, however, confounds the positive autocorrelation in the covariate with a positive state dependence.

We report the simulation results of our experiment with standard normal error terms in Figures 3 and 4.[8] Over all scenarios, the behavior of the tests based on models QE2 and QE1 and of Halliday's tests do not change remarkably when the distribution of ε_{it} is normal, except for a slight tendency to over-reject the true null hypothesis in "QE2_cov" and in "QE1_cov" when $\beta = 1$.

Figures 5 and 6 show the simulation results for a design where x_{it} follows a static process. When $\beta = 0$, the five tests perform in the same way as in a design with an autocorrelated covariate (see Figure 1). In contrast, when $\beta = 1$ and with $T = 5$, the proposed test "QE2_cov"maintains good size and power properties as well as the test based on model QE1 (albeit exhibiting less power), while Halliday's test and the tests "QE2_nocov" and "QE1_nocov" greatly over-reject the true null hypothesis.

Finally, Figures 7 and 8 show the results of the experiment design with predetermined covariates outlined at the end of Section 4.1 where the covariates follow the process defined in (20) with $\rho = 0.5$ and $\eta = 0.5$. The test statistic computed by means of the

[8]Simulation results for the upper and lower tailed tests are available upon request.

two-step procedure described in Section 3.4 is labeled "QE2_pred". The behavior of the test based on the two-step QE2 in the case of $v_i = 0$ (Figure 7) mostly remains unchanged, with a slight loss of power for small values of T. However, Figure 8 shows that, in the presence of unobserved heterogeneity in the model for the predetermined covariate, "QE2_pred" tends to slightly over-reject the null hypothesis when $\gamma = 0$, especially with T large. Notwithstanding, the original version of the test "QE2_cov" does not exhibit a reliable behavior, as it over-rejects the true null hypothesis far too often with both large n and T.

In conclusion, the simulation results show that the t-test based on model QE2 has good size and power properties that improve when n and T increase and are overall robust when the distribution of the time-varying error term is misspecified and in the presence of predetermined covariates, if the two-step version of QE2 is implemented.

The comparison with the test based on model QE1 and with Halliday's test confirm the superiority of the proposed t-test. As expected, the power properties of the test based on the initial QE1 model are less satisfactory compared to those of the proposed test. This is due to an information loss: model QE1 only considers the information concerning pairs of consecutive observations such that $(y_{i,t-1} = y_{it} = 1)$. The above simulation results also confirm our conjecture that, in absence of individual covariates and when $T = 2$, the proposed t-test for state dependence performs similarly, in terms of size and power, to the test proposed by Halliday (2007). Furthermore, in all other situations, our test is superior to the other, mainly due to the fact that its nominal size level is attained with any $T \geq 2$ and in the presence of individual covariates.

5. EMPIRICAL APPLICATION

In this section, we illustrate the proposed test by means of two examples. The first one is based on the PSID dataset and analyzes state dependence in female labor force participation and fertility. The purpose of this empirical application is to verify that results given by the proposed test are coherent with those found in the literature. It is well known that the individual's employment history shows positive state dependence (see, for instance, Hyslop, 1999; Carrasco, 2001) while fertility is negatively serially correlated (Bartolucci and Farcomeni, 2009).

In the second example, we test for state dependence in SRHS using data form the HRS dataset. In general, health outcome variables exhibit a high level of persistence. Such a persistence, however, can be due to both time-invariant unobserved heterogeneity and state dependence. The proposed test allows us to detect the state dependence effect while properly taking into account individual time-invariant effects. Since health variables receive considerable attention within health care policy, empirical studies have recently started to analyze the persistence in SRHS: Heiss (2011) finds that SRHS is positively autocorrelated in HRS; Halliday (2008) uses PSID data; Carro and Traferri (2012) analyze the British Household Panel Survey.

5.1. Example 1: Employment and Fertility

We base our example on a dataset derived from the Panel Study of Income Dynamics. Our application is closely related to the empirical analyses in Hyslop (1999) and in Bartolucci and Farcomeni (2009) that focus on the effect of fertility on women employment and on the magnitude of the state dependence effect in both variables. The dataset concerns $n = 1446$ married women between 18 and 46 years of age followed for $T = 5$ time occasions.

We compute the proposed t-test statistic for $H_0 : \gamma = 0$ based on model QE2 defined in (11) and compare it with Halliday's test statistic defined in (5) for each triple $(y_{i,t-1}, y_{it}, y_{i,t+1})$, $t = 1, \ldots, T-1$, applying the Bonferroni corrected nominal size (see Section 4). Since several time-varying individual characteristics are available, we compute the test based on model QE2 using the following covariates: number of children in the family between 1 and 2 years of age "child 1-2" and, similarly, "child 3-5", "child 6-14", "child 14-", "income" of the husband in dollars, time-dummies (1987 is taken as initial condition), "lagged employment", and "lagged fertility". In modeling employment as a function of fertility, a possible endogeneity problem may arise as the labor force and fertility decisions may be jointly determined (see Carrasco, 2001, and references therein). Therefore, we also compute the test based on the two-step estimation of model QE2 (see Section 3.4), where the first step consists of a regression of "fertility" on an intercept, "lagged employment", and "lagged fertility". Similarly, we compute the test using the two-step QE2 when modeling fertility as a function of employment.

If the null hypothesis of no state dependence is rejected, it is necessary to estimate the model parameters in a dynamic framework. As illustrated in Sections 2.1 and 2.2, suitable approaches are based on the CML estimator of Honoré and Kyriazidou (2000), the PCML estimator (Bartolucci and Nigro, 2012), or biased-corrected fixed-effects estimators (Carro, 2007). If the null hypothesis is not rejected, we simply estimate a static logit model.

For the dataset here considered, Table 2 reports the test statistics using the complete sample of $n = 1446$ women. The proposed test, indicated by QE2_cov, strongly rejects the null hypothesis of absence of state dependence for both response variables employment and fertility. The same result is obtained with the test based on the two-step QE2 model, indicated by QE2_pred. The signs of the test statistics also indicate positive state dependence for employment and negative for fertility. The same happens for the Halliday's test statistics, the values of which lead to reject the null hypothesis: in both cases there is at least one p–value lower than the Bonferroni corrected nominal size.

Since the null hypothesis of absence of state dependence is rejected, we estimated the dynamic logit model by the PCML estimator. The estimation results, reported in Table 3, confirm a strong positive state dependence for employment with an estimated coefficient $\hat{\gamma}$ close to 1.550 and a negative state dependence for fertility with $\hat{\gamma}$ equal to -0.906.

We further expand our application by analyzing the role of education in the labor force participation decision and fertility. The information available is the number of years of schooling in 1987 so, since time–invariant effects cancel out in model QE2, the test must be performed separately for different values of the educational attainment. For our exercise, we grouped observations of women with at most 12 years of schooling and those of women with more than 12 years of schooling (12 years corresponds to complete compulsory education).

Table 4 shows the test statistics for the two groups: in both cases the positive state dependence in employment is detected by all the test statistics, as they reject H_0. On the contrary, Halliday's test does not seem to detect the negative state dependence in fertility for less educated women; this is an important difference with respect to our approach. For this sub-sample, the PCML estimate of the state dependence parameter is -0.145 but not significantly different form 0. Overall, in this case the higher power of the proposed testing approach emerges over both the Halliday's approach and the approach based on the PCML estimator. Finally, for women with more than 12 years of schooling, there is agreement between all the tests as these reject the null hypothesis of absence of state dependence in fertility. For this sub-sample, the PCML estimate of γ is equal to -1.275.

5.2. Example 2: Self Reported Health Status

In the following, we test for state dependence in SRHS using a dataset based on HRS. Moreover, we further extend the analysis by analyzing the persistence in the health condition within different socio-demographic groups, separated by gender, race and education. In fact, along with SRHS, the dataset contains information on the individual's age at the time of the interview, gender, race, and education. After dropping observations with missing data in the variables of interest, we end up with a dataset of $n = 7074$ individuals for $T = 8$ periods, evenly interviewed form 1992 to 2006.

The response variable SRHS contains ordered responses to the question "How is your health in general?": it takes values 1 to 5 for the categories "excellent", "very good", "good", "fair", and "poor". For more accurate descriptive statistics, we refer the reader to Bartolucci et al. (2014a). In order to perform the proposed test for state dependence, we dichotomized SRHS: we tested for state dependence using all possible dichotomizations of the response variable generating the dummy variables $y_j = 1\,(\text{SRHS} > j)$ for $j = 1, \ldots, 4$. We performed the proposed test by estimating model QE2 which includes the age of the individual as covariate. Gender, race (white and non-white) and education (less than college, some college, college and above) are, instead, time–invariant so we need to test for state dependence separately for socio-demographic subgroups.

Table 5 reports the results of the test for state dependence for the whole sample and the different subgroups.

We find that, considering all the individuals in the sample, there is a strong positive state dependence effect in the health status in all the dichotomizations considered and the same result is found for the different subgroups considered. Coherently with the results found by Carro and Traferri (2012) and Halliday (2008), the persistence in the response to the SRHS question persists beyond the unobserved individual fixed effect.

We further disaggregated our sample by taking male and female with different levels of educational attainment and within these groups we consider white and non–white individuals. The second part of Table 5 shows the results of the proposed test for these sub-groups. Both males and females with different educational levels display a strong positive state dependence over all the dichotomizations considered. The same results are found when we take white individuals separately, while for non white-individuals the persistence in the response variable is somewhat weaker: for non-white less educated females, we accept the null hypothesis of absence of state dependence in poor health status (y_4), and for all non-white individuals college educated or above there is not a strong state dependence effect for all the considered dichotomizations.

In conclusion, we find that SRHS exhibits a positive state dependence considering both the whole sample and different sub-groups. This effect seems to weaken when we consider state dependence in the extreme categories (poor and excellent health), for non-white and highly educated individuals.

6. CONCLUSIONS

In this paper, we propose a test for state dependence under the dynamic logit model with individual covariates. The test is based on a modified version of the quadratic exponential model proposed in Bartolucci and Nigro (2010) in order to exploit more information about the association between the response variables. We show that this model correctly identifies the presence of state dependence regardless of whether individual covariates are present or not. We also show how the test may be used, in a modified version, with predetermined covariates.

Our test directly compares with the one proposed by Halliday (2007), which, however, cannot be easily applied in a panel with more than two periods (further to the initial observation) and does not allow for individual covariates. In the special case of two time periods and no covariates, the proposed test employs the same information on the response variables as Halliday's.

We studied the finite–sample properties of the t-test for state dependence proposed in this paper by means of a comprehensive Monte Carlo experiment in which it is compared with the test proposed by Halliday (2007). Simulation results show that the proposed test attains the nominal size even with not large samples (500 sample units), while it exhibits satisfactory power properties with large sample sizes. As expected, ignoring the presence of time-varying covariates in testing for state dependence leads to mistakenly detect a significant persistence in the response variable: the proposed test maintains its size properties, whereas Halliday's test over-rejects the true null hypothesis of absence of state dependence. Moreover, when state dependence is negative and the covariate positively affects the response variable, Halliday's test shows a remarkable power loss.

This result is confirmed by our empirical study based on a dataset derived from the Panel Study of Income Dynamics: when using either the whole sample or different sub-samples, the proposed test always rejects the null hypothesis of absence of state dependence in fertility for parameter estimates of about -1, whereas Halliday's test may fail to detect the state dependence. We also performed the proposed test on the self reported health status based on a dataset derived from the Health and Retirement Study: we find that there is a strong positive state dependence effect in the health outcome variable that persists across gender, ethnicity, and different educational levels but not for highly educated non-white individuals.

Overall, the main advantages of the proposed test are the simplicity of use and its flexibility. In fact, it can be very simply performed and does not require to formulate any parametric assumption on the distribution of the individual-specific intercepts (or on the correlation between these intercepts and the covariates) as random-effects approaches instead require. Moreover, it may be used even with only two time occasions (further to an initial observations) and with individual covariates, including time-dummies.

Finally, it is worth noting that the proposed test, being based on a modified quadratic exponential model, is more powerful than a t-test test based on more traditional quadratic exponential models or on the PCML estimator of Bartolucci and Nigro (2012). This aspect also emerges from the simulation study and the empirical applications.

ACKNOWLEDGMENTS

Francesco Bartolucci and Claudia Pigini acknowledge the financial support from the grant RBFR12SHVV of the Italian Government (FIRB project "Mixture and latent variable models for causal inference and analysis of socio-economic data"). The collection of the PSID data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and the National Science Foundation under award number 1157698. The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

REFERENCES

Bartolucci, F., Bacci, S., Pennoni, F. (2014a). Longitudinal analysis of self-reported health status by mixture latent auto-regressive models. *Journal of the Royal Statistical Society - Series C* 63:267–288.

Bartolucci, F., Bellio, R., Sartori, N., Salvan, A. (2014b). Modified profile likelihood for fixed-effects panel data models. *Econometric Reviews*, in press.

Bartolucci, F., Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association* 104:816–831.

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.

Bettin, G., Lucchetti, R. (2012). Intertemporal remittance behaviour by immigrants in germany. Working Papers 385, Università Politecnica delle Marche (I), Dipartimento di Scienze Economiche e Sociali.

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

Brown, S., Ghosh, P., Taylor, K. (2012). The existence and persistence of household financial hardship. Working Papers 22, The University of Sheffield, Department of Economics.

Cappellari, L., Jenkins, S. P. (2004). Modelling low income transitions. *Journal of Applied Econometrics* 19:593–610.

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.

Carro, J. M., Traferri, A. (2012). State dependence and heterogeneity in health using a bias-corrected fixed-effects estimator. *Journal of Applied Econometrics* 29:181–207.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In: Heckman, J. J., Singer, B., eds. *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press.

Chamberlain, G. (1993). Feedback in panel data models. Technical report, Department of Economics, Harvard University.

Feller, W. (1943). On a general class of "contagious" distributions. *Annals of Mathematical Statistics* 1:389–400.

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

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

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.

Halliday, T. J. (2007). Testing for state dependence with time-variant transition probabilities. *Econometric Reviews* 26:685–703.

Halliday, T. J. (2008). Heterogeneity, state dependence and health. *The Econometrics Journal* 11:499–516.

Heckman, J. (1981). Heterogeneity and state dependence. In: Rosen, S., ed. *Studies in Labor Markets*. Chicago: Chicago University Press, pp. 91–140.

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

Heiss, F. (2011). Dynamics of self-rated health and selective mortality. *Empirical Economics* 40:119–140.

Hochberg, Y., Tamhane, A. (1987). *Multiple Comparison Procedures*. Wiley.

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

Hsiao, C. (2005). *Analysis of Panel Data*. 2nd ed. New York: Cambridge University Press.

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

Murphy, K. M., Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics* 3:370–379.

Pigini, C., Presbitero, A. F., Zazzaro, A. (2014). State Dependence in Access to Credit. Mo.Fi.R. Working Papers 102, Money and Finance Research group (Mo.Fi.R.) - Univ. Politecnica Marche - Dept. Economic and Social Sciences.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–26.

Table 1: Simulation Results for the t-Test (QE1 and QE2) and Halliday's Test Statistics: Bidirectional, $\rho = 0.5$, $\eta = 0$

	γ	*QE2_cov*	*QE1_cov*	*Halliday*	*QE2_nocov*	*QE1_nocov*
				$n = 500$		
$\beta = 0$	−1.0	0.854	0.709	0.846	0.855	0.710
$T = 2$	−0.5	0.313	0.195	0.312	0.314	0.191
	0.0	0.048	0.045	0.052	0.044	0.050
	0.5	0.261	0.130	0.270	0.262	0.130
	1.0	0.684	0.268	0.682	0.690	0.268
$\beta = 0$	−1.0	1.000	0.999	0.990	1.000	0.999
$T = 5$	−0.5	0.983	0.960	0.524	0.982	0.959
	0.0	0.054	0.046	0.053	0.055	0.047
	0.5	0.971	0.915	0.413	0.972	0.916
	1.0	1.000	0.999	0.929	1.000	0.999
$\beta = 1$	−1.0	0.523	0.347	0.122	0.122	0.102
$T = 2$	−0.5	0.158	0.110	0.064	0.064	0.058
	0.0	0.052	0.039	0.233	0.233	0.148
	0.5	0.137	0.091	0.583	0.583	0.311
	1.0	0.368	0.158	0.832	0.832	0.448
$\beta = 1$	−1.0	1.000	0.999	0.452	0.530	0.465
$T = 5$	−0.5	0.810	0.718	0.073	0.123	0.116
	0.0	0.056	0.050	0.283	0.951	0.875
	0.5	0.761	0.675	0.748	1.000	0.999
	1.0	1.000	0.992	0.981	1.000	0.999
				$n = 1000$		
$\beta = 0$	−1.0	0.989	0.935	0.987	0.988	0.933
$T = 2$	−0.5	0.557	0.351	0.556	0.550	0.354
	0.0	0.054	0.050	0.059	0.058	0.056
	0.5	0.474	0.220	0.474	0.471	0.222
	1.0	0.943	0.483	0.930	0.946	0.480
$\beta = 0$	−1.0	1.000	0.999	1.000	1.000	0.999
$T = 5$	−0.5	0.998	0.997	0.849	0.998	0.997
	0.0	0.060	0.053	0.058	0.061	0.051
	0.5	0.998	0.996	0.719	0.998	0.996
	1.0	1.000	0.999	1.000	1.000	0.999
$\beta = 1$	−1.0	0.815	0.600	0.171	0.172	0.149
$T = 2$	−0.5	0.270	0.169	0.078	0.076	0.062
	0.0	0.051	0.045	0.441	0.430	0.246
	0.5	0.243	0.136	0.850	0.848	0.525
	1.0	0.658	0.308	0.980	0.983	0.772
$\beta = 1$	−1.0	1.000	0.999	0.745	0.800	0.743
$T = 5$	−0.5	0.980	0.956	0.113	0.146	0.122
	0.0	0.043	0.044	0.417	1.000	0.995
	0.5	0.971	0.933	0.980	1.000	0.999
	1.0	1.000	0.999	1.000	1.000	0.999

("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included. ε_{it} is distributed as a standard logistic r.v.)

Table 2: Tests for State Dependence ($H_1 : \gamma \neq 0$): Proposed t-Test QE2 and Halliday's Test Statistics for the Overall PSID Dataset

	Employment		*Fertility*	
	t-stat.	*p-value*	*t-stat.*	*p-value*
QE2_cov	13.58	0.00	−6.80	0.00
QE2_pred	15.55	0.00	−7.93	0.00
Halliday's test				
S_1 (1st triple)	5.75	0.00	−4.74	0.00
S_2 (2nd triple)	4.80	0.00	−4.97	0.00
S_3 (3rd triple)	4.02	0.00	−1.09	0.27
S_4 (4th triple)	4.27	0.00	−5.10	0.00
Sample size	1446		1446	

Model QE2 is estimated with covariates; Bonferroni corrected nominal size: 0.010206; Results of the first step estimation for computing QE2_pred are (standard errors in parentheses):

- Employment equation: $\text{fert}_{it} = \underset{(0.005)}{0.070} - \text{fert}_{i,t-1}\,\underset{(0.011)}{0.027} - \text{empl}_{i,t-1}\,\underset{(0.006)}{0.003}$; $\hat{\tau} = \underset{(0.003)}{0.249}$
- Fertility equation: $\text{empl}_{it} = \underset{(0.008)}{0.267} + \text{empl}_{i,t-1}\,\underset{(0.009)}{0.622} - \text{fert}_{i,t-1}\,\underset{(0.016)}{0.054}$; $\hat{\tau} = \underset{(0.004)}{0.360}$

Table 3: Estimation Results Based on the PCML Approach (Bartolucci and Nigro, 2012): Overall PSID Dataset

	Employment				*Fertility*			
	coeff.	*s.e.*	*t-stat.*	*p-value*	*coeff.*	*s.e.*	*t-stat.*	*p-value*
Child 1–2	−0.675	0.13	−5.10	0.00	−0.719	0.15	−4.72	0.00
Child 3–5	−0.312	0.12	−2.52	0.01	−1.085	0.21	−5.05	0.00
Child 6–13	−0.032	0.12	−0.25	0.40	−1.055	0.26	−4.08	0.00
Child 14–	−0.010	0.14	−0.07	0.47	−0.800	0.43	−1.86	0.03
Income/1000	−0.007	0.00	−1.68	0.05	−0.000	0.00	−0.13	0.45
1989	0.089	0.14	−1.12	0.13	0.402	0.15	4.64	0.00
1990	0.317	0.13	0.65	0.26	0.445	0.19	2.64	0.00
1991	0.089	0.13	2.49	0.01	0.397	0.24	2.31	0.01
1992	0.001	0.13	0.67	0.25	0.448	0.29	1.66	0.05
Lag fertility	−0.185	0.17	−1.09	0.28	−0.906	0.21	−4.35	0.00
Lag employment	1.550	0.11	13.93	0.00	0.801	0.17	1.56	0.06

Table 4: Tests for State Dependence ($H_1 : \gamma \neq 0$): Proposed t-Test QE2 and Halliday's Test Statistics for the PSID Dataset, by Education

	Years of schooling ≤ 12				*Years of schooling* > 12			
	Employment		*Fertility*		*Employment*		*Fertility*	
	t-stat.	*p-value*	*t-stat.*	*p-value*	*t-stat.*	*p-value*	*t-stat.*	*p-value*
QE2_cov	10.18	0.00	−2.70	0.01	9.04	0.00	−6.25	0.00
QE2_pred	11.11	0.00	−4.07	0.00	10.90	0.00	−6.82	0.00
Halliday's test								
S_1 (1st triple)	3.04	0.00	−1.60	0.11	5.27	0.00	−5.79	0.00
S_2 (2nd triple)	3.49	0.00	−0.81	0.41	3.30	0.00	−6.28	0.00
S_3 (3rd triple)	2.83	0.00	−0.79	0.43	2.75	0.01	−0.78	0.00
S_4 (4th triple)	4.40	0.00	−1.18	0.24	1.68	0.09	−5.45	0.00
Sample size		773				673		

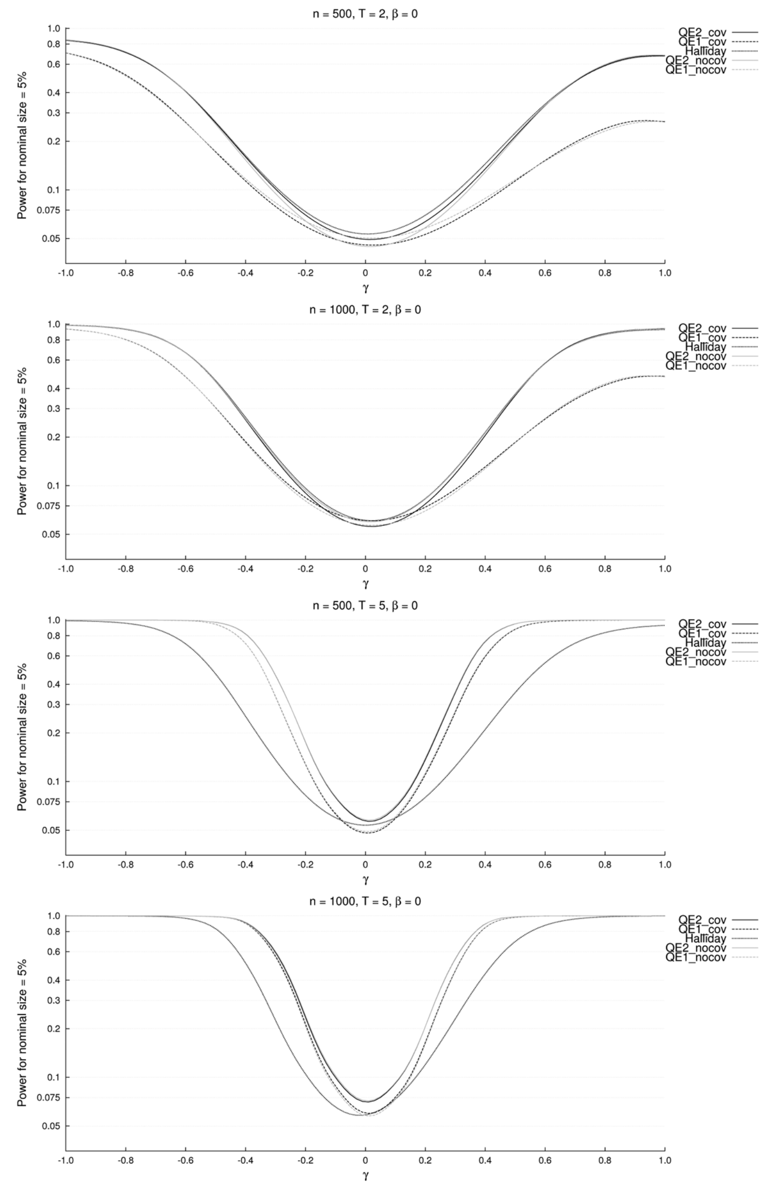
Model QE2 is estimated with covariates; Bonferroni corrected nominal size: 0.010206; Results of the first step estimation for computing QE2_pred are (standard errors in parentheses):

- Employment equation (sch. $\leq$ 12): $\text{fert}_{it} = \underset{(0.006)}{0.049} + \text{fert}_{i,t-1}\,\underset{(0.014)}{0.001} - \text{empl}_{i,t-1}\,\underset{(0.007)}{0.000}$; $\hat{\tau} = \underset{(0.003)}{0.215}$
- Fertility equation (sch. $\leq$ 12): $\text{empl}_{it} = \underset{(0.010)}{0.243} + \text{empl}_{i,t-1}\,\underset{(0.012)}{0.639} - \text{fert}_{i,t-1}\,\underset{(0.024)}{0.020}$; $\hat{\tau} = \underset{(0.005)}{0.363}$
- Employment equation (sch. $>$ 12): $\text{fert}_{it} = \underset{(0.009)}{0.101} - \text{fert}_{i,t-1}\,\underset{(0.016)}{0.059} - \text{empl}_{i,t-1}\,\underset{(0.010)}{0.011}$; $\hat{\tau} = \underset{(0.004)}{0.281}$
- Fertility equation (sch. $>$ 12): $\text{empl}_{it} = \underset{(0.012)}{0.299} + \text{empl}_{i,t-1}\,\underset{(0.014)}{0.599} - \text{fert}_{i,t-1}\,\underset{(0.020)}{0.087}$; $\hat{\tau} = \underset{(0.006)}{0.357}$

The PCML estimates are available upon request.

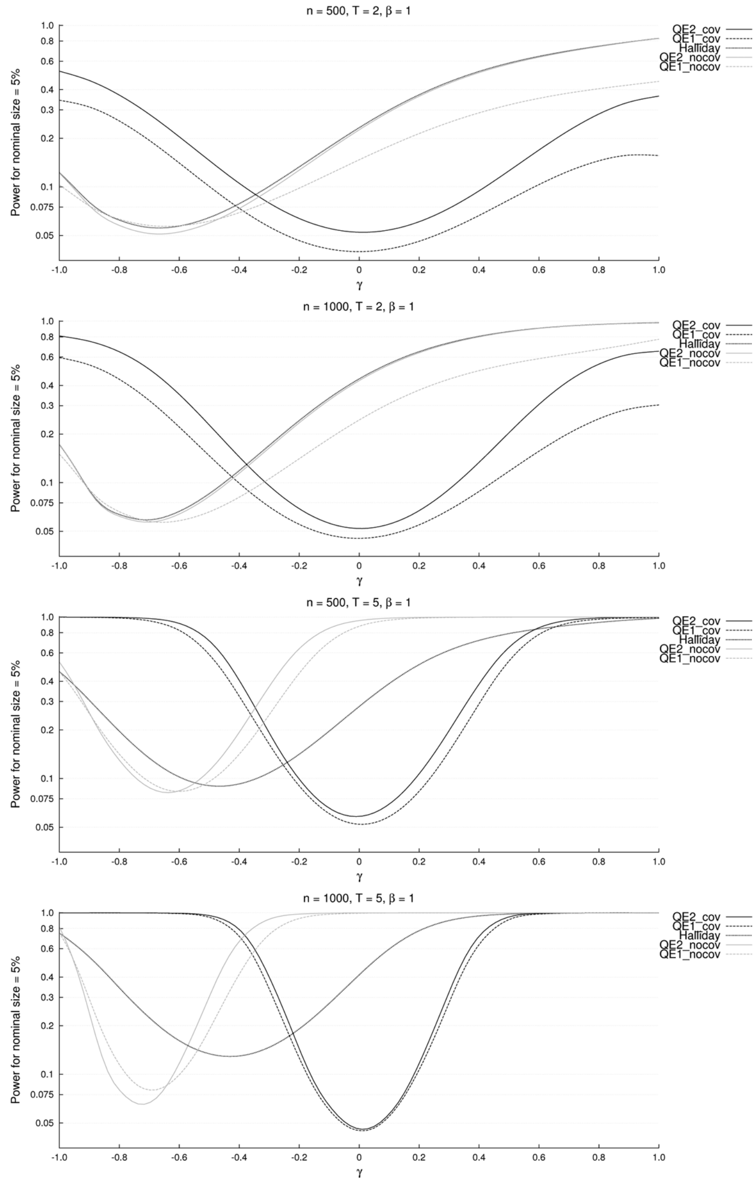
Table 5: Tests for State Dependence ($H_1 : \gamma \neq 0$): Proposed t-Test QE2 for the RAND HRS Dataset

		y_1		y_2		y_3		y_4	
	s. size	*t-test*	*p-value*	*t-test*	*p-value*	*t-test*	*p-value*	*t-test*	*p-value*
Total	7,074	15.17	.000	20.73	.000	20.07	.000	12.91	.000
Male	2,697	10.08	.000	13.53	.000	13.20	.000	9.77	.000
Female	4,107	11.38	.000	15.69	.000	15.09	.000	8.58	.000
White	5,863	14.30	.000	19.03	.000	18.40	.000	12.52	.000
Non-white	1,211	5.65	.000	8.40	.000	7.85	.000	3.40	.001
Less than c.	4,309	13.37	.000	17.56	.000	14.49	.000	8.53	.000
Some c.	1,395	5.93	.000	9.47	.000	9.95	.000	6.67	.000
C. and above	1,370	3.43	.001	5.36	.000	9.30	.000	6.68	.000
Male, Less than college									
White	1,398	8.55	.000	10.55	.000	8.38	.000	5.52	.000
Non-white	312	2.34	.019	4.07	.000	4.29	.000	1.75	.081
All	1,710	8.49	.000	11.24	.000	9.45	.000	5.76	.000
Female, Less than college									
White	2,050	9.40	.000	12.65	.000	10.36	.000	6.60	.000
Non-white	549	4.35	.000	5.38	.000	3.42	.001	−0.04	.969
All	2,599	10.34	.000	13.52	.000	10.98	.000	6.30	.000
Male, Some college									
White	475	3.97	.000	6.02	.000	5.56	.000	3.76	.000
Non-white	68	1.53	.127	2.36	.018	2.82	.005	3.39	.001
All	543	4.18	.000	6.47	.000	6.17	.000	4.78	.000
Female, Some college									
White	727	3.70	.000	5.44	.000	6.95	.000	4.96	.000
Non-white	125	1.44	.149	4.59	.000	3.53	.000	−0.56	.577
All	852	3.97	.000	6.75	.000	7.73	.000	4.72	.000
Male, College and above									
White	662	2.11	.035	3.36	.001	6.60	.000	5.69	.000
Non-white	52	0.51	.607	1.06	.287	0.10	.921	1.96	.050
All	714	2.20	.028	3.55	.000	6.43	.000	6.00	.000
Female, College and above									
White	551	3.07	.000	3.86	.000	5.86	.000	3.24	.001
Non-white	105	−0.29	.772	1.13	.259	3.44	.001	0.65	.513
All	656	2.75	.006	4.08	.000	6.75	.000	3.33	.001



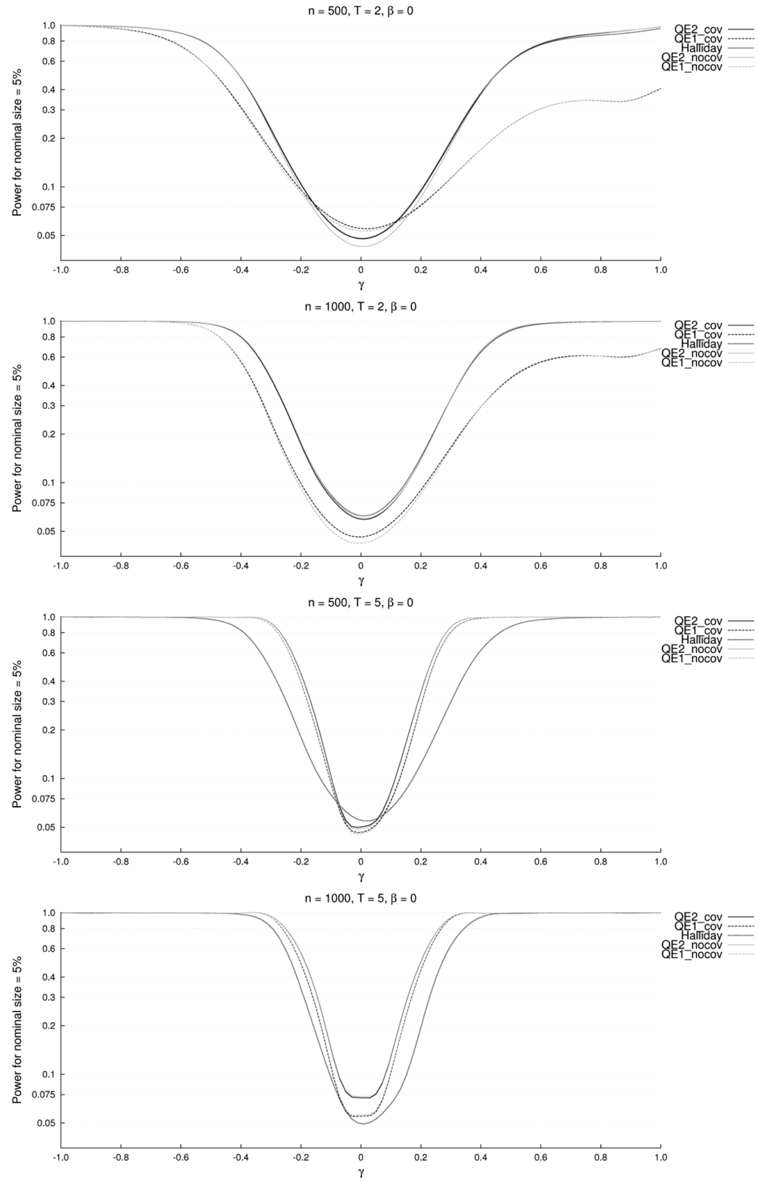
("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard logistic r.v. Results are plot in log-scale)

Figure 1: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), $\beta = 0$, $\rho = 0.5$, $\eta = 0$.



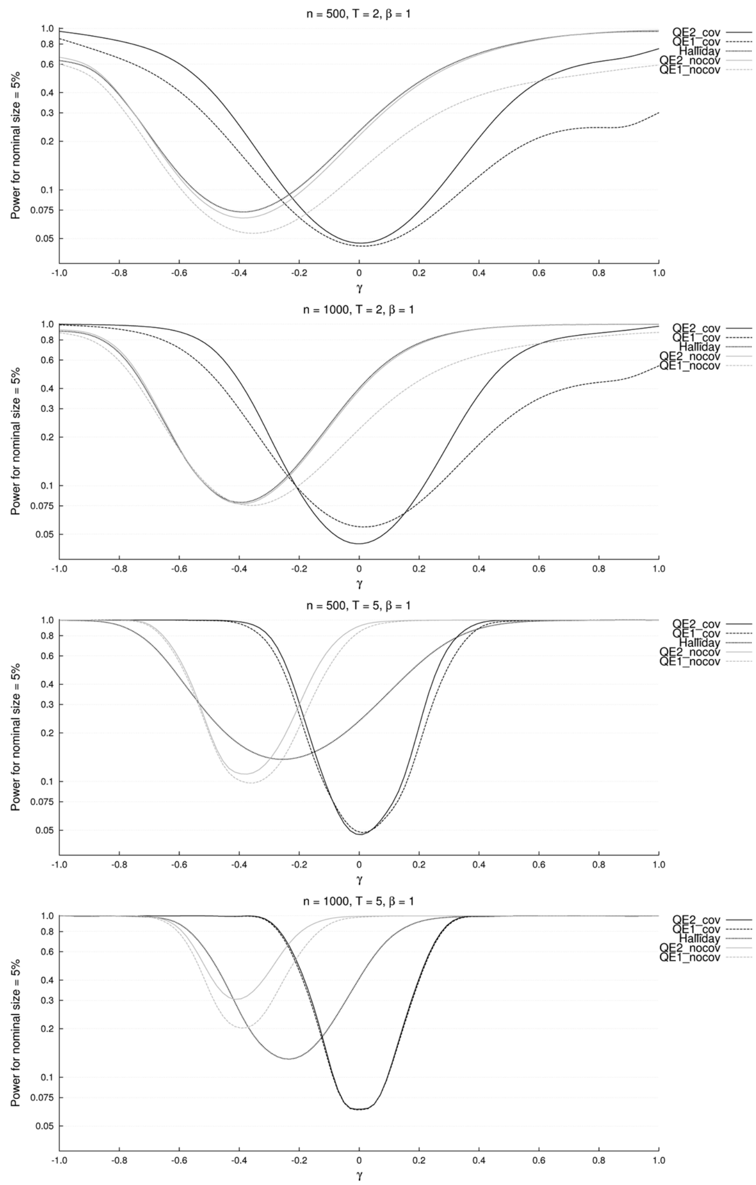
("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard logistic r.v. Results are plot in log-scale)

Figure 2: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), $\beta = 1$, $\rho = 0.5$, $\eta = 0$.



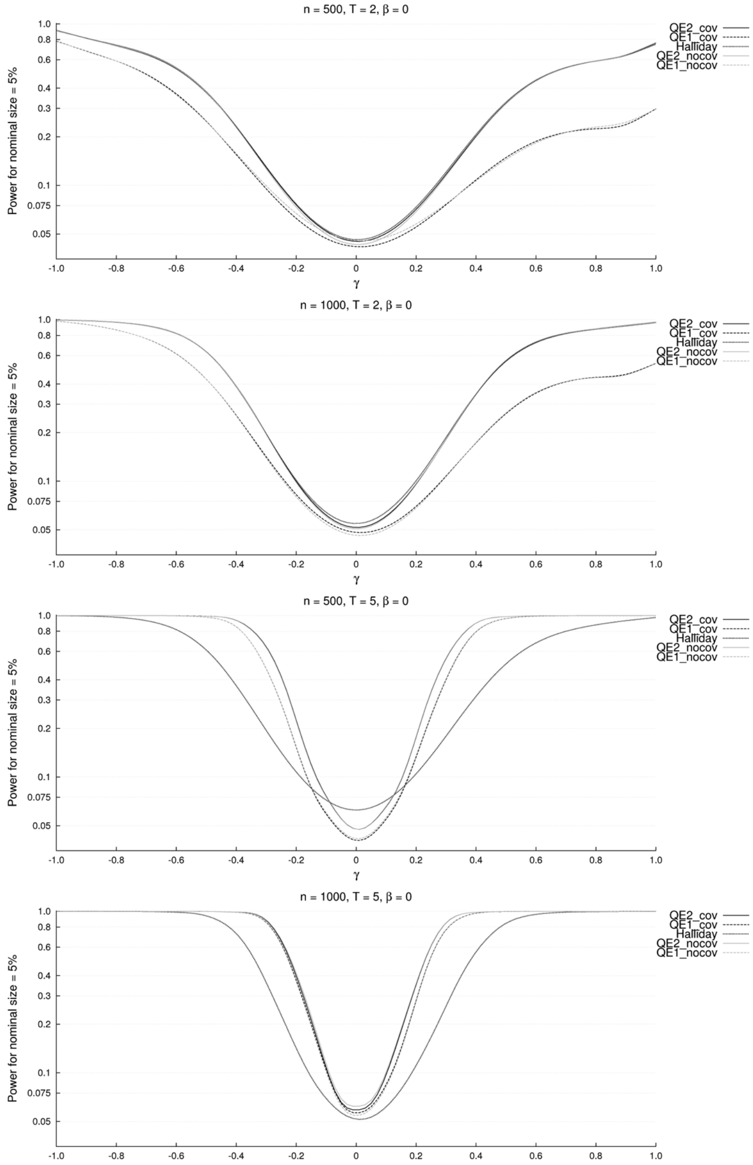
("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard normal r.v. Results are plot in log-scale)

Figure 3: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), normal error term, $\beta = 0$, $\rho = 0.5$, $\eta = 0$.



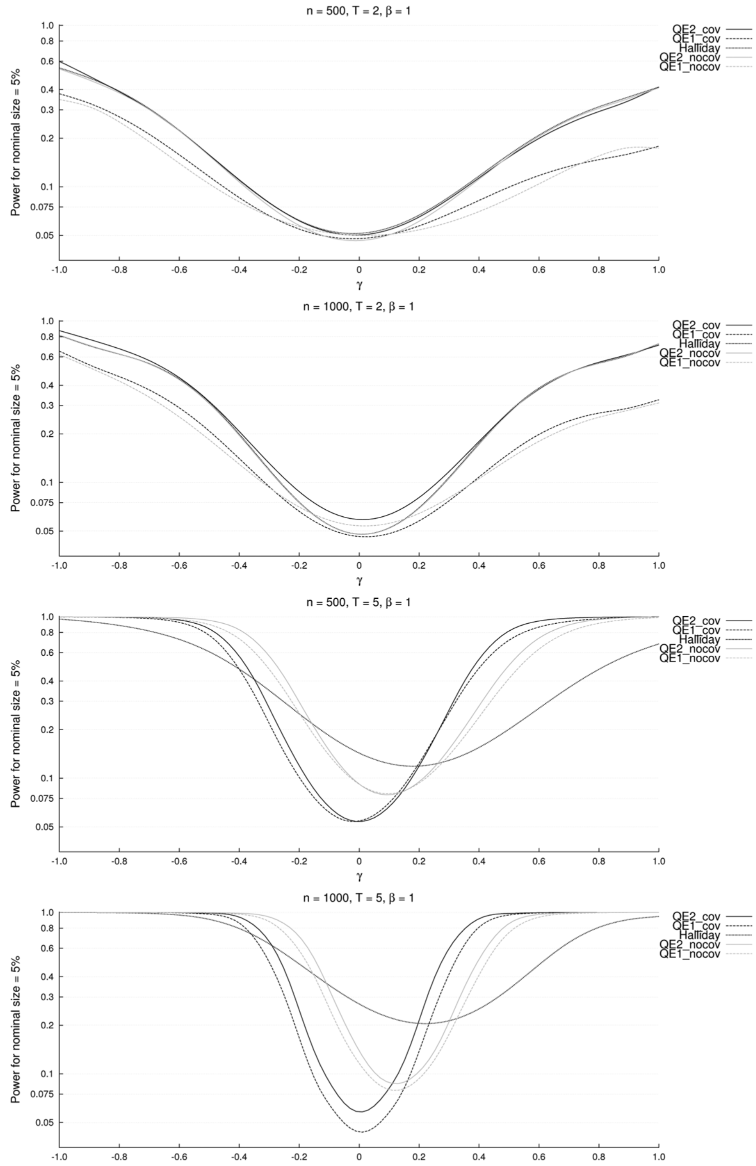
("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard normal r.v. Results are plot in log-scale)

Figure 4: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), normal error term, $\beta = 1$, $\rho = 0.5$, $\eta = 0$.



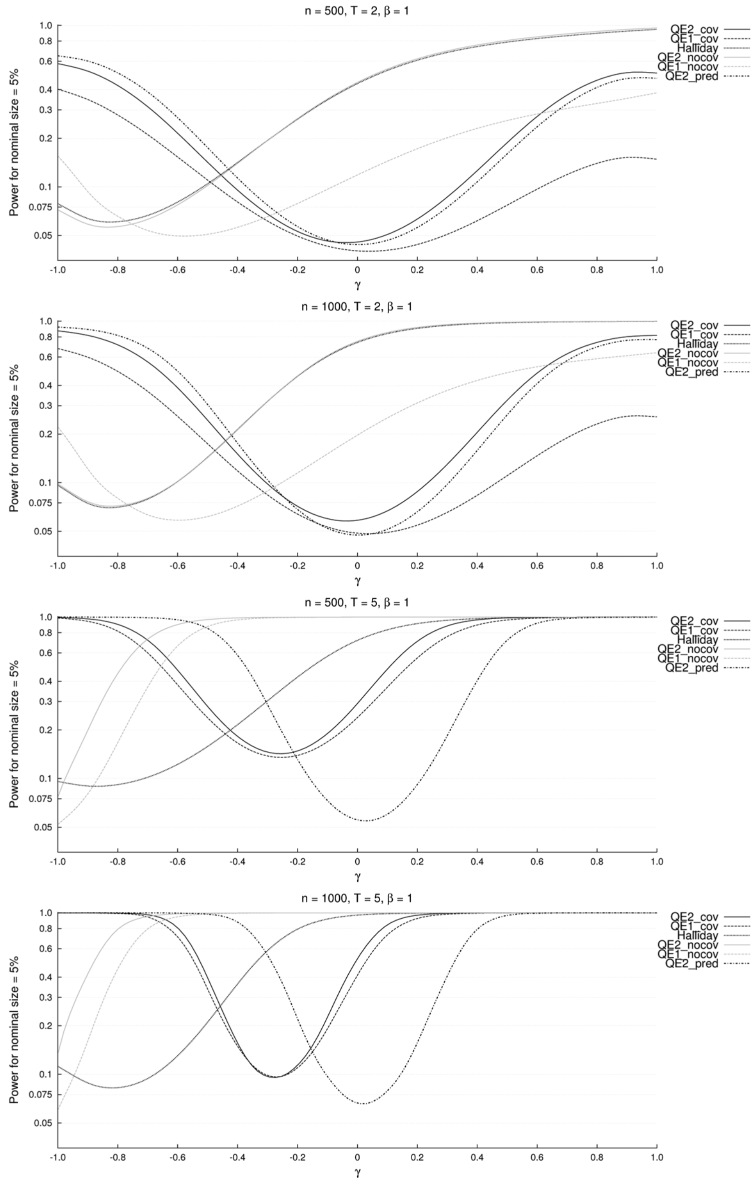
("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard logistic r.v. Results are plot in log-scale)

Figure 5: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), $\beta = 0$, $\rho = 0$, $\eta = 0$.



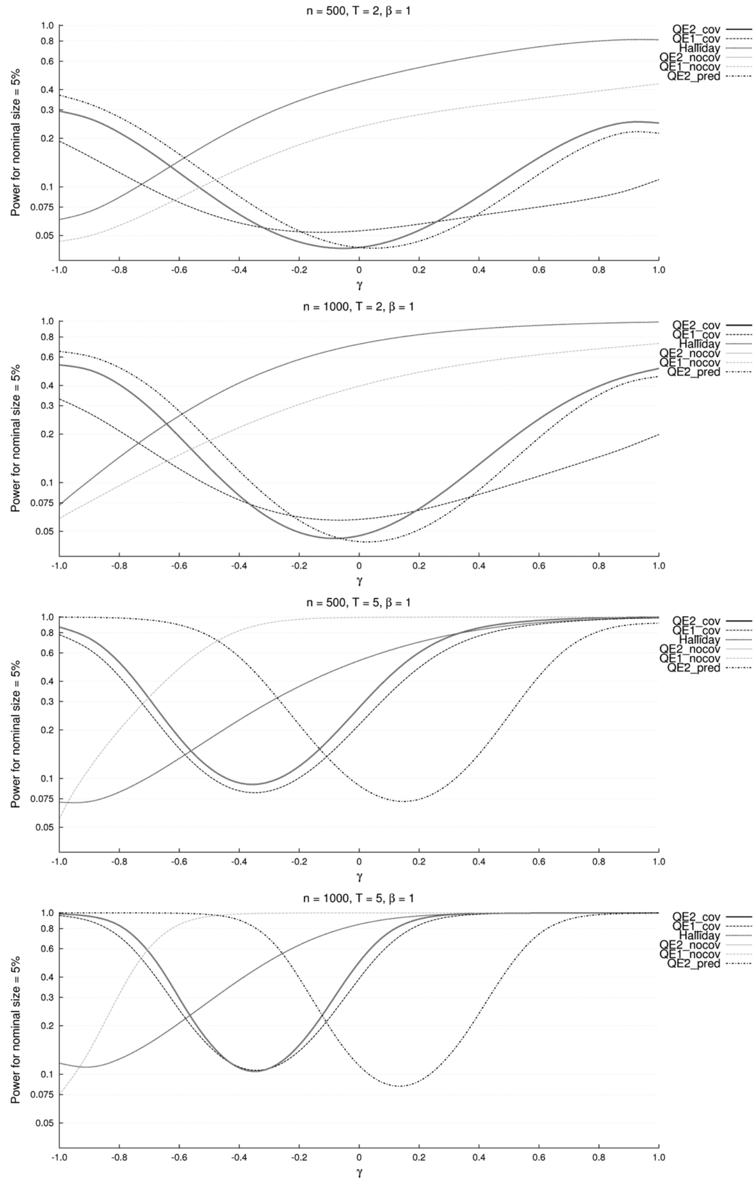
("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard logistic r.v. Results are plot in log-scale)

Figure 6: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), $\beta = 1$, $\rho = 0$, $\eta = 0$.



("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard logistic r.v. Results are plot in log-scale)

Figure 7: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), $\beta = 1$, $\rho = 0.5$, $\eta = 0.5$, $v_i = 0$, $i = 1, \ldots, n$ (predetermined covariates).



("QE1_cov" and "QE2_cov" refer to the case in which the covariate x_{it} is included in the QE1/QE2 model; "QE1_nocov" and "QE2_nocov" refer to the case in which the covariate is not included; power curves are smoothed by means of cubic splines. ε_{it} is distributed as a standard logistic r.v. Results are plot in log-scale)

Figure 8: Power plots for the t-test (QE2 and QE1) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$), $\beta = 1$, $\rho = 0.5$, $\eta = 0.5$, $v_i \sim N(0, 1)$, $i = 1, \ldots, n$ (predetermined covariates).

Journal of Econometrics 170 (2012) 102–116

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data[☆]

Francesco Bartolucci [a,*], Valentina Nigro [b]

[a] Dipartimento di Economia, Finanza e Statistica, Università di Perugia, 06123 Perugia, Italy

[b] Banca d'Italia, Via Nazionale 91, 00184 Roma, Italy

ARTICLE INFO

Article history:
Received 16 December 2009
Received in revised form
21 November 2011
Accepted 29 March 2012
Available online 23 April 2012

JEL classification:
C13
C23
C25

Keywords:
Log-linear models
Longitudinal data
Pseudo likelihood inference
Quadratic exponential distribution

ABSTRACT

We show how the dynamic logit model for binary panel data may be approximated by a quadratic exponential model. Under the approximating model, simple sufficient statistics exist for the subject-specific parameters introduced to capture the unobserved heterogeneity between subjects. The latter must be distinguished from the state dependence which is accounted for by including the lagged response variable among the regressors. By conditioning on the sufficient statistics, we derive a pseudo conditional likelihood estimator of the structural parameters of the dynamic logit model, which is simple to compute. Asymptotic properties of this estimator are studied in detail. Simulation results show that the estimator is competitive in terms of efficiency with estimators recently proposed in the econometric literature.

1. Introduction

One of the most important econometric models for binary panel data is the dynamic logit model, which includes, among the regressors, individual-specific intercepts for the unobserved heterogeneity and the lagged response variable for the true state dependence (Feller, 1943; Heckman, 1981a,b); see Hsiao (2005) for a review and Bartolucci and Farcomeni (2009) for extended versions of this model.

The individual-specific intercepts, included in the dynamic logit model, may be treated as fixed or random parameters. The fixed-parameters approach has the advantage of not requiring the formulation of any distribution on these parameters and of naturally addressing the well-known problem of the initial conditions; see Heckman (1981c) and Wooldridge (2000). On the other hand, the dynamic logit model with fixed-effects suffers from the *incidental parameter* problem (Neyman and Scott, 1948) and then the standard maximum likelihood estimator of the parameters of interest, for the covariates and the state dependence, is not consistent as the sample size grows to infinity. Using the statistical terminology, they will be referred to as the *structural parameters*.

A well-known method to overcome the problem of the incidental parameters consists of conditioning the inference on suitable sufficient statistics for these parameters. When the lagged response variable is omitted from the model, and therefore true state dependence is ruled out, sufficient statistics for the incidental parameters are the sums of the response variables at individual level, which will be referred to as the *total scores* (see Rasch, 1961). The resulting maximum likelihood estimator of the structural parameters may be computed by a simple Newton–Raphson algorithm and has optimal asymptotic properties (see Andersen, 1970, 1972). A conditional likelihood approach can also be followed when the assumed logit model includes the lagged response variable. This approach was developed by Honoré and Kyriazidou (2000) who, by employing some results of Chamberlain (1985), proposed a weighted conditional likelihood estimator of the structural parameters. The sufficient statistics on which this approach is based are different from the total scores and are

☆ The authors are grateful to Prof. F. Peracchi for his comments and suggestions. F. Bartolucci acknowledges the financial support from the ‘‘Einaudi Institute for Economics and Finance’’ (EIEF), Rome (IT). Most of the article has been developed during the period spent by V. Nigro at the University of Rome ‘‘Tor Vergata’’ and is part of her Ph.D. dissertation. The views are personal and do not involve the responsibility of the institutions with which the authors are affiliated.

* Corresponding author. Tel.: +39 075 5855227.
E-mail addresses: bart@stat.unipg.it (F. Bartolucci), valentina.nigro@bancaditalia.it (V. Nigro).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.03.004

such that a larger number of response configurations does not contribute to the likelihood. Moreover, the approach requires the specification of a suitable kernel function for weighting the response configuration of each subject in the sample on the basis of the covariates, implying the exclusion of time dummies and the reduction of the rate of convergence of the estimator to the true parameter value.

An alternative to conditional likelihood estimators is represented by bias corrected estimators, which have a reduced order of bias without increasing the asymptotic variance (see Hahn and Newey, 2004; Carro, 2007; Fernandez-Val, 2009; Hahn and Kuersteiner, 2011). The main advantage of this method is its general applicability to other dynamic models, beyond the logit one. Another positive aspect is the possibility to estimate policy parameters which depend on the fixed-effects, but note that marginal effect estimators have reduced bias only for long panels.

In this paper, we propose a pseudo conditional likelihood approach for estimating the dynamic logit model, which is based on approximating it by a model of quadratic exponential type (Cox, 1972). The approximating model is very similar to that proposed by Bartolucci and Nigro (2010) and corresponds to a log-linear model for the conditional distribution of the response variables given the initial observation and the covariates. The two-way interaction effects of this model are equal to a common parameter when they are referred to a pair of consecutive response variables and to 0 otherwise; moreover, up to a correction term, the main effects directly depend on the covariates and on individual-specific parameters for the unobserved heterogeneity. We show that the interaction parameter may be interpreted as in the dynamic logit model in terms of *log-odds ratio*, a well-known measure of association between binary variables (Agresti, 2002, Ch. 8).

It is worth noting that, although the statistical literature sometimes criticizes the use of log-linear models for the analysis of binary longitudinal data (see Diggle et al., 2002; Molenberghs and Verbeke, 2004), Bartolucci and Nigro (2010) showed that the model they developed has a meaningful interpretation in terms of expectation about future outcomes. Moreover, as for the Rasch (1961) model, the total scores are sufficient statistics for the incidental parameters. Then, the structural parameters may be estimated by a conditional maximum likelihood estimator which is $\sqrt{n}$-consistent even in the presence of aggregate variables, which are time-specific and common to all the subjects, such as time dummies.

We also show how to construct a pseudo conditional likelihood estimator of the structural parameters of the dynamic logit model which is based on the quadratic exponential approximation of this model. The estimator is simple to compute and does not require to formulate a weighting function as the estimator of Honoré and Kyriazidou (2000) does. Moreover, its asymptotic properties are studied on the basis of standard inferential results on maximum likelihood estimation of misspecified models (White, 1982; Newey and McFadden, 1994). In particular, we show that the proposed estimator is consistent for the vector of *pseudo true parameters*; in absence of state dependence, this vector coincides with the true parameter vector. Finite sample properties of the proposed estimator are studied by a series of simulations performed along the same lines as in Honoré and Kyriazidou (2000) and Carro (2007). These simulations show that the estimator is usually more efficient than alternative estimators.

Finally, we outline some extensions of the proposed approach to the case of dynamic logit models including a second-order lagged response variable and to that of categorical response variables with more than two categories. Note that the approach could also be adopted to estimate the dynamic probit model that, together with the dynamic logit model, is a workhorse model for binary panel data. In this way, we can reach a level of generality similar to that of the approach of Carro (2007).

The paper is organized as follows. In the next section we briefly review the relevant literature for the proposed approach. The approximating model used within this approach is described in Section 3, where its conditional distribution given the total scores is also derived. The resulting pseudo conditional maximum likelihood estimator is proposed in Section 4. Moreover, in Section 5 we illustrate the asymptotic properties, under the true logit model, of this estimator and in Section 6 we show the results of the simulation study. Finally, in Section 7 we outline some possible extensions of the proposed approach and in Section 8 we draw the main conclusions.

All the algorithms described in this paper have been implemented in MATLAB functions which are available from the webpage www.stat.unipg.it/~bart.

2. Preliminaries

With reference to a sample of n subjects observed at T consecutive occasions, let y_{it} be the binary random variable for subject i at occasion (or period) t, with $i = 1, \dots, n$ and $t = 1, \dots, T$, and let $\boldsymbol{x}_{it}$ be a corresponding vector of exogenous observable covariates. In the following, we first review the dynamic logit model for data of this type and the methods of Honoré and Kyriazidou (2000) and Carro (2007) for the estimation of its parameters. We then review the quadratic exponential model of Bartolucci and Nigro (2010) as a valid alternative to the dynamic logit model.

2.1. Dynamic logit model

In the econometric literature, binary data models are generally represented through a latent index function allowing for unobserved heterogeneity and first-order state dependence, that is

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

where $1\{\cdot\}$ is the indicator function, α_i is a fixed individual-specific parameter, ε_{it} represents the stochastic error term, and the initial observation y_{i0} is assumed to be exogenous. The parameters of primary interest are $\boldsymbol{\beta}$ and γ, which are the structural parameters and, in the following, will be jointly denoted by $\boldsymbol{\theta} = (\boldsymbol{\beta}', \gamma)'$. In particular, γ is the state dependence parameter which is assumed to be constant across individuals. The parameters α_i are instead considered as incidental parameters. Nevertheless, they cannot be omitted from the model in order to prevent biased estimation of the state dependence effect.

The dynamic logit model results when the errors terms ε_{it} are supposed independent and identically distributed, conditionally on the covariates and on the parameters α_i, with standard logistic distribution. Therefore, the conditional distribution of the overall vector of response variables $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})$ given α_i, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1} \cdots \boldsymbol{x}_{iT})$, and y_{i0} may be expressed as

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp\left(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}{}'\boldsymbol{\beta} + y_{i*}\gamma\right)}{\prod_t [1 + \exp(\alpha_i + \boldsymbol{x}_{it}{}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]}, \tag{2}$$

where $y_{i+} = \sum_t y_{it}$ and $y_{i*} = \sum_t y_{i,t-1}y_{it}$, with the product $\prod_t$ and the sum $\sum_t$ ranging over $t = 1, \dots, T$.

An interesting approach for estimating the fixed-effects model illustrated above is based on the maximization of the conditional likelihood given suitable statistics for the incidental parameters. In particular, Honoré and Kyriazidou (2000), extending the conditional approach of Chamberlain (1985), proposed an estimator based on the maximization of a weighted conditional log-likelihood. For $T = 3$, this log-likelihood is defined as follows

$$\sum_i 1\{y_{i1}+y_{i2}=1\}K\left(\frac{\boldsymbol{x}_{i2}-\boldsymbol{x}_{i3}}{\sigma_n}\right)\log[r(\boldsymbol{y}_i|\alpha_i,\boldsymbol{X}_i,y_{i0},\quad y_{i1}+y_{i2}=1,y_{i3},\boldsymbol{x}_{i2}=\boldsymbol{x}_{i3})], \tag{3}$$

where $K(\cdot)$ is a kernel function with bandwidth σ_n a priori fixed and

$$r(\boldsymbol{y}_i|\alpha_i,\boldsymbol{X}_i,y_{i0},y_{i1}+y_{i2}=1,y_{i3},\boldsymbol{x}_{i2}=\boldsymbol{x}_{i3}) = \frac{\exp\{y_{i1}[(\boldsymbol{x}_{i1}-\boldsymbol{x}_{i2})'\boldsymbol{\beta}+(y_{i0}-y_{i3})\gamma]\}}{1+\exp[(\boldsymbol{x}_{i1}-\boldsymbol{x}_{i2})'\boldsymbol{\beta}+(y_{i0}-y_{i3})\gamma]}.$$

Note that the weight given to the response configuration of subject i decreases with the distance between $\boldsymbol{x}_{i2}$ and $\boldsymbol{x}_{i3}$ and a large weight is given to the response configuration of this subject when $\boldsymbol{x}_{i2}$ is close to $\boldsymbol{x}_{i3}$, and then the property of conditional independence of $\boldsymbol{y}_i$ from α_i approximately holds.

The fixed-effects approach of Honoré and Kyriazidou (2000) has the advantage of not requiring particular assumptions either on the unobserved heterogeneity or on the initial conditions. However, its use requires a careful choice of the kernel function and of its bandwidth. This choice affects the rate of convergence of the estimator to the true parameter value. The rate of convergence is in any case slower than $\sqrt{n}$. Moreover, since only certain response configurations are considered (such that $y_{i1}+y_{i2} = 1$ and $\boldsymbol{x}_{i2}$ is near to $\boldsymbol{x}_{i3}$ with $T = 3$), the *actual sample size*[1] is usually much smaller than the nominal sample size n; this limits the efficiency of the estimator. Furthermore, aggregate variables are not identified in this approach because of the support condition required for the covariates. For further comments see Magnac (2004) and Honoré and Tamer (2006).

A recent field of research is based on a different approach to the estimation of the dynamic discrete choice models with fixed-effects, proposing bias corrected estimators. These estimators have a reduced order of bias with respect to the conventional maximum likelihood estimator, without having a higher asymptotic variance (see Hahn and Newey, 2004; Carro, 2007; Fernandez-Val, 2009; Hahn and Kuersteiner, 2011). In particular, Carro (2007) showed that the correction of the score function reduces the order (in T) of its bias from $O(1)$ to $O(T^{-1})$, giving an estimator unbiased to order $O(T^{-2})$. Although this estimator is only consistent when the number of time periods goes to infinity, Monte Carlo simulations have shown its good finite sample performance in comparison to the estimator of Honoré and Kyriazidou (2000), even with not very long panels (e.g., eight time periods).

2.2. Quadratic exponential model

The family of quadratic exponential models was firstly proposed by Cox (1972) for the analysis of multivariate binary data. Models belonging to this class are log-linear models in which all the effects of order higher than two are equal to zero. Their use for the analysis of binary longitudinal data has been already considered in the statistical literature; for a review see Diggle et al. (2002) and Molenberghs and Verbeke (2004).

Bartolucci and Nigro (2010) introduced a model belonging to the above family for which they provide a meaningful interpretation. The model assumes that the joint response probability for subject i is given by

$$p(\boldsymbol{y}_i|\delta_i,\boldsymbol{X}_i,y_{i0}) = \frac{\exp\left[y_{i+}\delta_i+\sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi}_1+y_{iT}(\psi+\boldsymbol{x}_{iT}'\boldsymbol{\phi}_2)+y_{i*}\tau\right]}{\sum_{\boldsymbol{z}}\exp\left[z_+\delta_i+\sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi}_1+z_T(\psi+\boldsymbol{x}_{iT}'\boldsymbol{\phi}_2)+z_{i*}\tau\right]}, \tag{4}$$

where the sum $\sum_{\boldsymbol{z}}$ ranges over all the possible binary response vectors $\boldsymbol{z}=(z_1,\dots,z_T)$; moreover, $z_+=\sum_t z_t$ and $z_{i*}=y_{i0}z_1+\sum_{t>1}z_{t-1}z_t$. This model closely resembles the dynamic logit model based on the joint probability (2) and, as such, it allows for state dependence and unobserved heterogeneity, beyond the effects of the available covariates. Note that, in order to avoid confusion with the dynamic logit model, we now denote the incidental parameters by δ_i, the parameter vectors for the covariates by $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, and the parameter for the state dependence by τ.

The model parameters may be interpreted by considering that assumption (4) implies that

$$p(y_{it}|\delta_i,\boldsymbol{X}_i,y_{i0},\dots,y_{i,t-1}) = \frac{\exp\{y_{it}[\delta_i+\boldsymbol{x}_{it}'\boldsymbol{\phi}_1+y_{i,t-1}\tau+e_t^*(\delta_i,\boldsymbol{X}_i)]\}}{1+\exp[\delta_i+\boldsymbol{x}_{it}'\boldsymbol{\phi}_1+y_{i,t-1}\tau+e_t^*(\delta_i,\boldsymbol{X}_i)]},$$

where, for $t < T$ we have

$$e_t^*(\delta_i,\boldsymbol{X}_i) = \log\frac{1+\exp[\delta_i+\boldsymbol{x}_{i,t+1}'\boldsymbol{\phi}_1+e_{t+1}^*(\delta_i,\boldsymbol{X}_i)+\tau]}{1+\exp[\delta_i+\boldsymbol{x}_{i,t+1}'\boldsymbol{\phi}_1+e_{t+1}^*(\delta_i,\boldsymbol{X}_i)]} = \log\frac{p(y_{i,t+1}=0|\delta_i,\boldsymbol{X}_i,y_{it}=0)}{p(y_{i,t+1}=0|\delta_i,\boldsymbol{X}_i,y_{it}=1)},$$

and

$$e_T^*(\delta_i,\boldsymbol{X}_i) = \psi+\boldsymbol{x}_{iT}'\boldsymbol{\phi}_2. \tag{5}$$

The last expression is a reduced form for the correction term for the last time period. This correction term depends on future covariates and it is therefore approximated by a linear form of the covariate vector $\boldsymbol{x}_{iT}$.

Even if the above model is here used as a tool for estimating the dynamic logit model, it is worth noting that it is equivalent to a latent index model with error terms logistically distributed and systematic part including a correction term $e_t^*(\delta_i,\boldsymbol{X}_i)$, besides the usual covariates. This term may be interpreted as a measure of the effect of the present choice y_{it} on the expected utility (or propensity) at the next occasion ($t+1$). Moreover, as under the dynamic logit, y_{it} is conditionally independent of any other response variable given $y_{i,t-1}$ and $y_{i,t+1}$ and the parameter τ for the state dependence is the log-odds ratio between any pair of variables $(y_{i,t-1},y_{it})$, conditional on all the other response variables or marginal with respect to these variables. For a more detailed description of these properties, which are related to the model interpretation, and in particular of Eq. (5), we refer to Bartolucci and Nigro (2010).

From the point of view of inference, the main advantage of the above model is that the parameters for the unobserved heterogeneity may be eliminated by conditioning on the sums of the response variables across time. In this way, the structural parameters are identified with at least two observations further to the initial observation ($T \geq 2$), even in the presence of time dummies, giving a $\sqrt{n}$-consistent estimator. This estimator is computed by means of a simple Newton–Raphson algorithm which maximizes the log-likelihood based on the conditional probability

$$p(\boldsymbol{y}_i|\delta_i,\boldsymbol{X}_i,y_{i0},y_{i+}) = \frac{\exp\left[\sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi}_1+y_{iT}(\psi+\boldsymbol{x}_{iT}'\boldsymbol{\phi}_2)+y_{i*}\tau\right]}{\sum_{\boldsymbol{z}:z_+=y_{i+}}\exp\left[\sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi}_1+z_T(\psi+\boldsymbol{x}_{iT}'\boldsymbol{\phi}_2)+z_{i*}\tau\right]}, \tag{6}$$

[1] The actual sample size is the number of response configurations which contribute to the likelihood.

where the sum $\sum_{\boldsymbol{z}:z_+=y_{i+}}$ is extended to all response configurations $\boldsymbol{z}$ with sum equal to y_{i+}.

The absence of assumptions on the support of the covariates implies a larger actual sample exploited by this estimator with respect to that of Honoré and Kyriazidou (2000), and then a higher efficiency.

3. Proposed approximation

In this section, we propose an approximation of the dynamic logit model illustrated in Section 2.1 through a quadratic exponential model. We also discuss the main features of the approximating model in comparison to the true model.

3.1. Approximating quadratic exponential model

Along the same lines followed by Cox and Wermuth (1994), Bartolucci and Pennoni (2007), and Bartolucci (2010) in different contexts, we first take the logarithm of the joint probability $p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0})$ as defined in (2) under the dynamic logit model, that is

$$\log[p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0})] = y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}{}'\boldsymbol{\beta} + y_{i*}\gamma - \sum_t \log[1+\exp(\alpha_i + \boldsymbol{x}_{it}{}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]. \quad (7)$$

Then, we approximate the component which is not linear in the parameters through a first-order Taylor-series expansion around $\alpha_i = \bar{\alpha}_i$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, and $\gamma = 0$, obtaining[2]

$$\sum_t \log[1+\exp(\alpha_i + \boldsymbol{x}_{it}{}'\boldsymbol{\beta} + y_{i,t-1}\gamma)] \approx \sum_t \{\log[1+\exp(\bar{\alpha}_i + \boldsymbol{x}_{it}{}'\bar{\boldsymbol{\beta}})] + \bar{q}_{it}[\alpha_i - \bar{\alpha}_i + \boldsymbol{x}_{it}{}'(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})]\} + \bar{q}_{i1}y_{i0}\gamma + \sum_{t>1} \bar{q}_{it}y_{i,t-1}\gamma, \quad (8)$$

where $\bar{\alpha}_i$ and $\bar{\boldsymbol{\beta}}$ denote fixed values of α_i and $\boldsymbol{\beta}$, respectively, and

$$\bar{q}_{it} = \frac{\exp(\bar{\alpha}_i + \boldsymbol{x}_{it}{}'\bar{\boldsymbol{\beta}})}{1+\exp(\bar{\alpha}_i + \boldsymbol{x}_{it}{}'\bar{\boldsymbol{\beta}})}. \quad (9)$$

The last one is the expression of the probability that $y_{it} = 1$ when the parameters are fixed as above. Note that only the last sum at the rhs of expression (8) depends on the response configuration $\boldsymbol{y}_i$. Therefore, by substituting (8) in (7) and renormalizing the exponential of the resulting expression, we obtain the approximation

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) \approx p^*(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}),$$

with

$$p^*(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp\left(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}{}'\boldsymbol{\beta} - \sum_{t>1}\bar{q}_{it}y_{i,t-1}\gamma + y_{i*}\gamma\right)}{\sum_{\boldsymbol{z}}\exp\left(z_+\alpha_i + \sum_t z_t\boldsymbol{x}_{it}{}'\boldsymbol{\beta} - \sum_{t>1}\bar{q}_{it}z_{t-1}\gamma + z_{i*}\gamma\right)}, \quad (10)$$

with $\sum_{\boldsymbol{z}}$ and z_{i*} defined as in (4). In applying this approximation to estimate the parameter of the dynamic logit model, the terms $\bar{q}_{it}$ will be chosen in a suitable way.

From expression (10), we easily recognize that the approximating model is a modified version of the quadratic exponential model of Bartolucci and Nigro (2010), illustrated in Section 2.2. Moreover, the joint probability under the approximating model mimics expression (2) which holds under the true dynamic logit model, the main difference being in the denominator which in (10) does not depend on $\boldsymbol{y}_i$ and is simply a *normalizing constant* that may be denoted by $\mu(\alpha_i, \boldsymbol{X}_i, y_{i0})$. Also note that the true model and the approximating model coincide when there is no state dependence, both of them reducing to the static logit model. In fact, with $\gamma = 0$ we have

$$p^*(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp\left(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}{}'\boldsymbol{\beta}\right)}{\sum_{\boldsymbol{z}}\exp\left(z_+\alpha_i + \sum_t z_t\boldsymbol{x}_{it}{}'\boldsymbol{\beta}\right)} = \prod_t \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}_{it}{}'\boldsymbol{\beta})]}{1+\exp(\alpha_i + \boldsymbol{x}_{it}{}'\boldsymbol{\beta})}, \quad (11)$$

which does not depend either on $\bar{\alpha}_i$ or $\bar{\boldsymbol{\beta}}$.

The strong connection between the true model and the approximating model is clarified in the following Theorem, which may be proved along the same lines as in Bartolucci and Nigro (2010).

Theorem 1. *For $i = 1, \ldots, n$, quadratic exponential model (10) implies that the conditional logit of y_{it}, given α_i, $\boldsymbol{X}_i$, and $y_{i0}, \ldots, y_{i,t-1}$, is equal to*

$$\log\frac{p^*(y_{it}=1|\alpha_i, \boldsymbol{X}_i, y_{i0}, \ldots, y_{i,t-1})}{p^*(y_{it}=0|\alpha_i, \boldsymbol{X}_i, y_{i0}, \ldots, y_{i,t-1})} = \begin{cases} \alpha_i + \boldsymbol{x}_{it}{}'\boldsymbol{\beta} + y_{i,t-1}\gamma + e_t(\alpha_i, \boldsymbol{X}_i) - \bar{q}_{i,t+1}\gamma, & \text{if } t < T, \\ \alpha_i + \boldsymbol{x}_{it}{}'\boldsymbol{\beta} + y_{i,t-1}\gamma, & \text{if } t = T, \end{cases} \quad (12)$$

where

$$e_t(\alpha_i, \boldsymbol{X}_i) = \log\frac{p^*(y_{i,t+1}=0|\alpha_i, \boldsymbol{X}_i, y_{it}=0)}{p^*(y_{i,t+1}=0|\alpha_i, \boldsymbol{X}_i, y_{it}=1)}.$$

This correction term depends on the data only through $\boldsymbol{x}_{i,t+1}, \ldots, \boldsymbol{x}_{i,T}$ and is such that $e_t(\alpha_i, \boldsymbol{X}_i) \approx \bar{q}_{i,t+1}\gamma$, $t = 2, \ldots, T$, where the approximation is in the sense of (8).

For $i = 1, \ldots, n$, model (10) also implies that:

(i) *y_{it} is conditionally independent of $y_{i0}, \ldots, y_{i,t-2}$ given α_i, $\boldsymbol{X}_i$, and $y_{i,t-1}$ $(t = 2, \ldots, T)$;*
(ii) *y_{it} is conditionally independent of $y_{i0}, \ldots, y_{i,t-2}, y_{i,t+2}, \ldots, y_{iT}$, given α_i, $\boldsymbol{X}_i$, $y_{i,t-1}$, and $y_{i,t+1}$ $(t = 2, \ldots, T-1)$.*

Note that, for $t = T$, expression (12) is based exactly on the same parametrization adopted under the dynamic logit model. When $t < T$, this equivalence holds approximately since $e_t(\alpha_i, \boldsymbol{X}_i) \approx \bar{q}_{i,t+1}\gamma$. The above Theorem also implies that

$$\log\frac{p^*(y_{it}=1|\alpha_i, \boldsymbol{X}_i, y_{i,t-1}=1)}{p^*(y_{it}=0|\alpha_i, \boldsymbol{X}_i, y_{i,t-1}=1)} - \log\frac{p^*(y_{it}=1|\alpha_i, \boldsymbol{X}_i, y_{i,t-1}=0)}{p^*(y_{it}=0|\alpha_i, \boldsymbol{X}_i, y_{i,t-1}=0)} = \gamma,$$
$$i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

and then, under the approximating model, γ may be interpreted as the log-odds ratio between any consecutive pair of response variables, conditional on or marginal with respect to all the other response variables. This is the same interpretation that γ has under the dynamic logit. Moreover, the approximating model reproduces the same conditional independence relations between the response variables (see results (i) and (ii) above) of the dynamic logit model.

---

[2] As for the quality of approximation, from standard results on Taylor-series expansions we have that the remainder term R is bounded above as follows:

$$R \le 0.25\sum_t \{(\alpha_i - \bar{\alpha}_i)^2/2 + [\boldsymbol{x}_{it}{}'(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})]^2/2 + y_{i,t-1}\gamma^2/2 + (\alpha_i - \bar{\alpha}_i)\boldsymbol{x}_{it}{}'(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) + y_{i,t-1}(\alpha_i - \bar{\alpha}_i)\gamma + y_{i,t-1}\boldsymbol{x}_{it}{}'(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\gamma\}.$$

3.2. Conditional distribution given the sufficient statistics

Regardless of the distribution of the covariates, the approximating model has minimal sufficient statistics for the heterogeneity parameters α_i, which are the total scores y_{i+}, $i = 1, \ldots, n$. The availability of these sufficient statistics is the main advantage with respect to the true model. In particular, expression (10) implies that the conditional distribution of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$, y_{i0}, and y_{i+} is

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

where the sum at the denominator is defined as in (6); this expression does not depend on α_i. Dividing the numerator and the denominator by $\exp(y_{i+}\boldsymbol{x}_{i1}'\boldsymbol{\beta})$, it may be reformulated in a simpler way as

$$p^*(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp\left(\sum_{t>1} y_{it}\boldsymbol{d}_{it}'\boldsymbol{\beta} - \sum_{t>1}\bar{q}_{it}y_{i,t-1}\gamma + y_{i*}\gamma\right)}{\sum_{\boldsymbol{z}:z_+=y_{i+}}\exp\left(\sum_{t>1} z_t\boldsymbol{d}_{it}'\boldsymbol{\beta} - \sum_{t>1}\bar{q}_{it}z_{t-1}\gamma + z_{i*}\gamma\right)}, \tag{13}$$

with $\boldsymbol{d}_{it} = \boldsymbol{x}_{it} - \boldsymbol{x}_{i1}$. Then, time-invariant covariates and the individual intercepts α_i are not identified. The same happens for all conditional approaches, such as that of Honoré and Kyriazidou (2000) and that employed by Bartolucci and Nigro (2010) to make inference on the quadratic exponential model. Moreover, as in the approach of Honoré and Kyriazidou (2000), we assume the strictly exogeneity of the regressors, which is a standard condition for the consistency of conditional likelihood estimators.

On the basis of an estimate for the structural parameters, each parameter α_i may be estimated by maximizing the corresponding log-likelihood under the true model, that is $\log p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0})$, by a standard algorithm. With short panels, more stable estimates of these parameters may be obtained by maximizing a modified version of this log-likelihood, which is formulated as proposed in McCullagh and Tibshirani (1990) or Firth (1993). The estimates of the individual intercepts α_i obtained in this way allow us to derive marginal effects in an obvious way. At this regard see also Fernandez-Val (2009).

A natural question that arises at this point is why we rely on a Taylor-series expansion around a point of the parameter space at which $\gamma = 0$, instead of considering a generic point $\alpha_i = \bar{\alpha}_i$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, $\gamma = \bar{\gamma}$. The first reason for doing this is that an expansion about $\gamma = \bar{\gamma}$ would result in a model that, although rather similar to that based on (10), has sufficient statistics for the incidental parameters which differ from the total scores and imposes too many restrictions on the support of the covariates. On the other hand, a series of simulations has shown that the estimator of $\boldsymbol{\theta}$ obtained by maximizing the pseudo conditional likelihood based on approximation (10) has a very low bias even when samples are generated from a dynamic logit model of type (2), in which the parameter γ is far from 0. See Section 6 for a detailed illustration of the results of these simulations.

4. Pseudo conditional likelihood estimator

In this Section, we introduce the pseudo conditional likelihood estimator based on the approximating model described above, which may be computed on the basis of an observed sample of size n, represented by $(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)$, with $i = 1, \ldots, n$.

4.1. Definition of the estimator

It is clear that the use of the approximating model, having joint probability mass function defined in (10), requires to fix the probabilities $\bar{q}_{it}$, $i = 1, \ldots, n$, $t = 2, \ldots, T$. At this aim, we rely on a preliminary estimation of the vector of the regression parameters for the covariates. Therefore, the proposed estimator of $\boldsymbol{\theta}$ is based on the following two steps:

1. Compute a preliminary estimate $\tilde{\boldsymbol{\beta}}$ of the vector of regression parameters by maximizing the conditional likelihood of the static logit model. We write this log-likelihood as
$$\ell(\bar{\boldsymbol{\beta}}) = \sum_i 1\{0 < y_{i+} < T\}\ell_i(\bar{\boldsymbol{\beta}}),$$
$$\ell_i(\bar{\boldsymbol{\beta}}) = \log\frac{\exp\left(\sum_{t>1} y_{it}\boldsymbol{d}_{it}'\bar{\boldsymbol{\beta}}\right)}{\sum_{\boldsymbol{z}:z_+=y_{i+}}\exp\left(\sum_{t>1} z_t\boldsymbol{d}_{it}'\bar{\boldsymbol{\beta}}\right)}, \tag{14}$$
which is the same conditional log-likelihood of the approximating model under $\gamma = 0$ and may be maximized by a standard Newton–Raphson algorithm. Note that we include $1\{0 < y_{i+} < T\}$ in the above expression because $\ell_i(\bar{\boldsymbol{\beta}})$ is equal to 0 for y_{i+} equal to 0 or T.
2. Estimate $\boldsymbol{\theta}$ by maximizing the conditional log-likelihood of the approximating model, based on (13), which has expression
$$\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \sum_i 1\{0 < y_{i+} < T\}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}),$$
$$\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \log[p^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+})]; \tag{15}$$
we add the subscript $\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}$ to $p^*(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+})$ in order to underline its dependence on $\boldsymbol{\theta}$ and on $\bar{\boldsymbol{\beta}}$ through the probabilities $\bar{q}_{it}$, $t = 2, \ldots, T$, with $\bar{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$. These probabilities are computed, for every i such that $0 < y_{i+} < T$, by Eq. (9), with each individual parameter $\bar{\alpha}_i$ equal to its maximum likelihood estimate under the same static logit model as above.[3]

The resulting *pseudo conditional likelihood* estimator is denoted by $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\gamma})'$.

Note that, even in expression (15) we include the indicator function $1\{0 < y_{i+} < T\}$, since $\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = 0$ when $y_{i+} = 0$ or $y_{i+} = T$. Then, the corresponding response configurations do not provide information on the parameters. The actual sample size is then smaller than the nominal one, but it is always larger than that we have in the approach of Honoré and Kyriazidou (2000), which is based on the weighted log-likelihood of type (3). With $T = 3$, for instance, the response configurations $\boldsymbol{y}_i$ omitted from (15) are $(0, 0, 0)$ and $(1, 1, 1)$, whereas the response configurations $(0, 0, 1)$ and $(1, 1, 0)$ are also omitted from (3).

In order to show how to maximize $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ and to study the properties of the proposed estimator $\hat{\boldsymbol{\theta}}$, it is convenient to express each component $\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ in the canonical exponential family form as

$$\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \boldsymbol{u}^*(y_{i0}, \boldsymbol{y}_i)'\boldsymbol{A}^*(\boldsymbol{X}_i)'\boldsymbol{\theta} - \log\sum_{\boldsymbol{z}:z_+=y_{i+}}\exp[\boldsymbol{u}^*(y_{i0}, \boldsymbol{z})'\boldsymbol{A}^*(\boldsymbol{X}_i)'\boldsymbol{\theta}], \tag{16}$$

[3] The maximum likelihood estimate of α_i is obtained by maximizing the individual log-likelihood

$$\sum_t \log\frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}})]}{1+\exp(\alpha_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}})} = \sum_t y_{it}(\alpha_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}}) - \log[1+\exp(\alpha_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}})].$$

The solution $\bar{\alpha}_i$ is simple to find and is such that $\sum_t \bar{q}_{it} = y_{i+}$.

with

$$\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)=\left(y_{i2},\ldots,y_{iT},y_{i*}-\sum_{t>1}\bar{q}_{it}y_{i,t-1}\right)'. \tag{17}$$

Moreover

$$\boldsymbol{A}^*(\boldsymbol{X}_i)=\begin{pmatrix}\boldsymbol{X}_i\boldsymbol{D}' & 0\\ \boldsymbol{0}' & 1\end{pmatrix}, \tag{18}$$

where $\boldsymbol{D}=\begin{pmatrix}-\boldsymbol{1} & \boldsymbol{I}\end{pmatrix}$, with $\boldsymbol{I}$ denoting an identity matrix of suitable dimension, is a matrix of contrasts such that $\boldsymbol{X}_i\boldsymbol{D}'=(\boldsymbol{d}_{i2}\cdots\boldsymbol{d}_{iT})$ and $\boldsymbol{0}$ denotes a column vector of zeros of suitable dimension. Consequently, the score vector $\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})=\nabla_{\boldsymbol{\theta}}\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ and the observed information matrix $\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})=-\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ may be found through standard results on the exponential family (Barndoff-Nielsen, 1978, Ch. 8). In particular, we have

$$\begin{aligned}\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) &= \sum_i 1\{0<y_{i+}<T\}\boldsymbol{A}^*(\boldsymbol{X}_i)\{\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)\\ &\quad -E^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)|\boldsymbol{X}_i,y_{i0},y_{i+}]\},\end{aligned} \tag{19}$$

$$\begin{aligned}\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) &= \sum_i 1\{0<y_{i+}<T\}\boldsymbol{A}^*(\boldsymbol{X}_i)\\ &\quad\times V^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)|\boldsymbol{X}_i,y_{i0},y_{i+}]\boldsymbol{A}^*(\boldsymbol{X}_i)',\end{aligned} \tag{20}$$

which depend on the following conditional expected value and variance

$$E^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)|\boldsymbol{X}_i,y_{i0},y_{i+}]=\sum_{\boldsymbol{z}:z_+=y_{i+}}\boldsymbol{u}^*(y_{i0},\boldsymbol{z})p^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}(\boldsymbol{z}|\boldsymbol{X}_i,y_{i0},y_{i+}),$$

$$\begin{aligned}&V^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)|\boldsymbol{X}_i,y_{i0},y_{i+}]\\ &\quad=E^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)'|\boldsymbol{X}_i,y_{i0},y_{i+}]\\ &\quad-E^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)|\boldsymbol{X}_i,y_{i0},y_{i+}]E^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)|\boldsymbol{X}_i,y_{i0},y_{i+}]'.\end{aligned}$$

Note that $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ is always concave since the observed information matrix $\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ is always non-negative definite, as it is the sum of a series of variance–covariance matrices. When the sample size is large enough, under identifiability conditions on the covariates (see Theorem 2 below), this matrix is almost surely positive definite. Therefore, $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ may be maximized by a simple Newton–Raphson algorithm. This algorithm performs a series of iterations until convergence. At the hth iteration, the estimate of $\boldsymbol{\theta}$ is updated as

$$\boldsymbol{\theta}^{(h)}=\boldsymbol{\theta}^{(h-1)}+\boldsymbol{J}^*(\boldsymbol{\theta}^{(h-1)}|\bar{\boldsymbol{\beta}})^{-1}\boldsymbol{s}^*(\boldsymbol{\theta}^{(h-1)}|\bar{\boldsymbol{\beta}}). \tag{21}$$

The estimate $\hat{\boldsymbol{\theta}}$ is then found at convergence of this algorithm. Usually the iterative algorithm rapidly converges to the maximum of $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$, given the concavity of this function.

How to obtain standard errors for the proposed estimator, taking even into account the first step required to choose $\bar{\boldsymbol{\beta}}$, is shown after an example based on a simple, but important, reference model.

4.2. Case of $T=2$ time occasions

In order to illustrate the proposed estimator, we consider the case of $T=2$ time-occasions with only time dummies. This example is closely related to that provided for the static logit model by Hsiao (2005, Sec. 7.3) and is based on the assumption

$$p(y_{i1}|\alpha_i,y_{i0})=\frac{\exp[y_{i1}(\alpha_i+y_{i0}\gamma)]}{1+\exp(\alpha_i+y_{i0}\gamma)},$$

$$p(y_{i2}|\alpha_i,y_{i1})=\frac{\exp[y_{i2}(\alpha_i+\beta+y_{i1}\gamma)]}{1+\exp(\alpha_i+\beta+y_{i1}\gamma)},$$

for $i=1,\ldots,n$. Note that this model has only two parameters, which are β, corresponding to the difference between the two regression coefficients for the two time dummies, and γ for the state dependence.

For the above model, at step 1 we compute the conditional estimator of β by the explicit formula

$$\tilde{\beta}=\log\frac{n_{001}+n_{101}}{n_{010}+n_{110}},$$

where $n_{y_0y_1y_2}$ denote the frequency of the response configuration (y_0,y_1,y_2). Then, for every i such that $y_{i+}=1$, at step 2 we compute $\bar{\alpha}_i=-\bar{\beta}/2$ and we let

$$\bar{q}_{i2}=\bar{q}_2=\frac{\exp(\bar{\beta}/2)}{1+\exp(\bar{\beta}/2)},$$

with $\bar{\beta}=\tilde{\beta}$. Moreover, we maximize the pseudo conditional log-likelihood (15), where each component $\ell^*_i(\boldsymbol{\theta}|\bar{\beta})$ is expressed as in (16), with $\boldsymbol{u}^*(y_{i0},\boldsymbol{y}_i)=(y_{i2},y_{i0}y_{i1}-\bar{q}_2y_{i1})'$ and $\boldsymbol{A}^*(\boldsymbol{X}_i)$ simply equal to an identity matrix of dimension 2. After some algebra, we have that

$$\ell^*(\boldsymbol{\theta}|\bar{\beta})=\sum_i 1\{y_{i+}=1\}[y_{i2}\beta+(y_{i0}y_{i1}-\bar{q}_2y_{i1})\gamma-\log k(y_{i0})],$$

$$k(y_{i0})=\exp[(y_{i0}-\bar{q}_2)\gamma]+\exp(\beta),$$

which may be also expressed as

$$\begin{aligned}\ell^*(\boldsymbol{\theta}|\bar{\beta}) &= (n_{001}+n_{101})\beta+[n_{110}-\bar{q}_2(n_{010}+n_{110})]\gamma\\ &\quad -m_0\log k(0)-m_1\log k(1),\end{aligned}$$

where $m_{y_0}=n_{y_001}+n_{y_010}$ is the frequency of the response configurations with initial observation equal to y_0 and $y_1\neq y_2$.

From (19), we have that the score is

$$\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\beta})=\sum_i 1\{y_{i+}=1\}\times\begin{pmatrix}y_{i2}-\dfrac{\exp(\beta)}{k(y_{i0})}\\ y_{i0}y_{i1}-\bar{q}_2y_{i1}-\dfrac{(y_{i0}-\bar{q}_2)\exp[(y_{i0}-\bar{q}_2)\gamma]}{k(y_{i0})}\end{pmatrix}.$$

Moreover, from (20) we have the following expression for the observed information matrix

$$\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\beta})=\sum_i\frac{1\{y_{i+}=1\}\exp[\beta+(y_{i0}-\bar{q}_2)\gamma]}{k(y_{i0})^2}\times\begin{pmatrix}1 & -(y_{i0}-\bar{q}_2)\\ -(y_{i0}-\bar{q}_2) & (y_{i0}-\bar{q}_2)^2\end{pmatrix}.$$

Using the frequencies $n_{y_0y_1y_2}$, we have the equivalent expressions which are given in Box I. It is worth noting that the determinant of the latter matrix is equal to

$$|\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\beta})|=\frac{\exp[2\beta+(1-2\bar{q}_2)\gamma]m_0m_1}{k(0)^2k(1)^2},$$

which is strictly positive if $m_0>0$ and $m_1>0$. Under this condition, the function to be maximized, $\ell^*(\boldsymbol{\theta}|\bar{\beta})$, is strictly concave and has only one maximum obtained by solving the equation $\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\beta})=\boldsymbol{0}$.

4.3. Standard errors

In order to derive an expression for the standard errors, we rely on the Generalized Method of Moments (GMM) approach (Hansen, 1982). In fact, the proposed estimation method consists of solving the score equation

$$\boldsymbol{g}(\bar{\boldsymbol{\beta}},\boldsymbol{\theta})=\sum_i 1\{0<y_{i+}<T\}\boldsymbol{g}_i(\bar{\boldsymbol{\beta}},\boldsymbol{\theta})=\boldsymbol{0},$$

$$\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\beta}) = \begin{pmatrix} n_{001} + n_{101} - \dfrac{m_0 \exp(\beta)}{k(0)} - \dfrac{m_1 \exp(\beta)}{k(1)} \\ n_{110} - \bar{q}_2(n_{010} + n_{110}) + \dfrac{m_0 \bar{q}_2 \exp(-\bar{q}_2\gamma)}{k(0)} - \dfrac{m_1(1-\bar{q}_2)\exp[(1-\bar{q}_2)\gamma]}{k(1)} \end{pmatrix} \tag{22}$$

and

$$\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\beta}) = \frac{m_0 \exp(\beta - \bar{q}_2\gamma)}{k(0)^2}\begin{pmatrix} 1 & \bar{q}_2 \\ \bar{q}_2 & \bar{q}_2^2 \end{pmatrix} + \frac{m_1 \exp[\beta + (1-\bar{q}_2)\gamma]}{k(1)^2}\begin{pmatrix} 1 & -(1-\bar{q}_2) \\ -(1-\bar{q}_2) & (1-\bar{q}_2)^2 \end{pmatrix}. \tag{23}$$

Box I.

where

$$\boldsymbol{g}_i(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta}) = \begin{pmatrix} \nabla_{\bar{\beta}} \ell_i(\bar{\boldsymbol{\beta}}) \\ \nabla_{\boldsymbol{\theta}} \ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}), \end{pmatrix}$$

with $\ell_i(\bar{\boldsymbol{\beta}})$ defined in (14) and $\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ defined in (15). The solution of this equation is represented by $(\tilde{\boldsymbol{\beta}}', \hat{\boldsymbol{\theta}}')'$.

Once the proposed method is casted into a GMM approach, we are legitimated to estimate the variance–covariance matrix of $(\tilde{\boldsymbol{\beta}}', \hat{\boldsymbol{\theta}}')'$ by

$$\boldsymbol{W}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \boldsymbol{H}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})^{-1}\boldsymbol{S}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})[\boldsymbol{H}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})^{-1}]', \tag{24}$$

where

$$\boldsymbol{S}(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\}\boldsymbol{g}_i(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta})\boldsymbol{g}_i(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta})'$$

and

$$\boldsymbol{H}(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\}\boldsymbol{H}_i(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta}),$$

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

is the derivative of $\boldsymbol{g}(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta})$ with respect to $(\bar{\boldsymbol{\beta}}', \boldsymbol{\theta}')$. In the above expression, $\boldsymbol{O}$ denotes a suitable matrix of zeros, whereas the expressions of the other blocks are given in Appendix.

Once the matrix $\boldsymbol{W}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ is computed as above, the standard errors for the pseudo conditional estimators in $\hat{\boldsymbol{\theta}}$ may be obtained in the usual way from the main diagonal of the lower right submatrix of $\boldsymbol{W}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$. Then, an approximate $(1-\alpha)$-level confidence interval may be constructed for any parameter β_h in $\boldsymbol{\beta}$ and for γ as follows

$$\hat{\beta}_h \mp z_{\alpha/2}\mathrm{se}(\hat{\beta}_h) \quad \text{and} \quad \hat{\gamma} \mp z_{\alpha/2}\mathrm{se}(\hat{\gamma}),$$

where se(·) denotes the standard error obtained as above and $z_{\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard Normal distribution.

5. Asymptotic properties of the pseudo conditional likelihood estimator

In this section, we deal with identifiability issues and asymptotic properties of the proposed estimator under the true model. At this regard, we assume that the data $(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)$, $i = 1, \ldots, n$, are independently drawn from the true model based on the density function $f_0(\boldsymbol{X}, y_0, \boldsymbol{y})$. The latter is obtained from the marginalization of

$$f_0(\alpha, \boldsymbol{X}, y_0, \boldsymbol{y}) = f_0(\alpha, \boldsymbol{X}, y_0)p_0(\boldsymbol{y}|\alpha, \boldsymbol{X}, y_0), \tag{25}$$

where $f_0(\alpha, \boldsymbol{X}, y_0)$ denotes the joint distribution of the individual-specific intercept α, the covariates $\boldsymbol{X} = (\boldsymbol{x}_1 \cdots \boldsymbol{x}_T)$, and the initial observation y_0. Moreover, $p_0(\boldsymbol{y}|\alpha, \boldsymbol{X}, y_0)$ denotes the conditional distribution of the response variables under dynamic logit model (2) when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, with $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \gamma_0)'$ denoting the true value of its structural parameters. By suitable marginalization of the densities in (25), we also obtain $f_0(\boldsymbol{X}, y_0)$, $p_0(\boldsymbol{y}|\boldsymbol{X}, y_0)$, and $f_0(\boldsymbol{X}, y_0, y_+)$, which will be used in the following.

Under the above assumption, we first investigate the issue of consistency of the proposed estimator $\hat{\boldsymbol{\theta}}$, which is strongly connected to that of the identification of the parameters (Newey and McFadden, 1994). Then, we deal with the asymptotic distribution of the estimator.

5.1. Consistency

Let $\bar{\boldsymbol{\beta}}_*$ the point at which the conditional estimator $\tilde{\boldsymbol{\beta}}$, computed at the step 1, converges in probability as n tends to infinity. In symbols, we have $\tilde{\boldsymbol{\beta}} \xrightarrow{p} \bar{\boldsymbol{\beta}}_*$ as $n \to \infty$. Then

$$\frac{\ell^*(\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}})}{n} \xrightarrow{p} E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)], \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{26}$$

where $\ell^*(\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}})$ is the pseudo conditional log-likelihood considered at step 2 and $\boldsymbol{\Theta}$ is the parameter space. Moreover, by $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$ we mean the expected value, under the true model, of the individual component of the log-likelihood defined in (16). More explicitly, we have that

$$E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)] = E_0\{E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)|\boldsymbol{X}, y_0]\}, \tag{27}$$

where the outer expected value at rhs is with respect to the distribution $f_0(\boldsymbol{X}, y_0)$, whereas the inner expected value is with respect to $p_0(\boldsymbol{y}|\boldsymbol{X}, y_0)$, that is

$$\begin{aligned} E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)|\boldsymbol{X}, y_0] = \sum_{\boldsymbol{y}} \Bigg\{ & \boldsymbol{u}^*(y_0, \boldsymbol{y})'\boldsymbol{A}^*(\boldsymbol{X})'\boldsymbol{\theta} \\ & - \log \sum_{\boldsymbol{z}: z_+ = y_+} \exp[\boldsymbol{u}^*(y_0, \boldsymbol{z})'\boldsymbol{A}^*(\boldsymbol{X})'\boldsymbol{\theta}] \Bigg\} \\ & \times p_0(\boldsymbol{y}|\boldsymbol{X}, y_0), \end{aligned} \tag{28}$$

with $\boldsymbol{u}^*(y_0, \boldsymbol{y})$ and $\boldsymbol{A}^*(\boldsymbol{X})$ defined in (17) and (18), respectively. It is important to recall that $\boldsymbol{u}^*(y_0, \boldsymbol{y})$ involves the probabilities $\bar{q}_{it}$ that, in computing (28), are substituted by

$$\bar{q}_{*t}(\boldsymbol{X}, \boldsymbol{y}) = \frac{\exp(\bar{\alpha}_* + \boldsymbol{x}_t'\bar{\boldsymbol{\beta}}_*)}{1 + \exp(\bar{\alpha}_* + \boldsymbol{x}_t'\bar{\boldsymbol{\beta}}_*)},$$

where $\bar{\alpha}_*$ is such that $\sum_t \bar{q}_{*t}(\boldsymbol{X}, \boldsymbol{y}) = y_+$.

A relevant aspect of $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$ is that it has first derivative corresponding to a single component of the sum used in (19) to define the score vector. This derivative is equal to $\boldsymbol{v}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$, with

$$\begin{aligned} \boldsymbol{v}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \nabla_{\boldsymbol{\theta}} E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})] = E_0(\boldsymbol{A}^*(\boldsymbol{X})\{ & \boldsymbol{u}^*(y_0, \boldsymbol{y}) \\ & - E^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_0, \boldsymbol{y})|\boldsymbol{X}, y_0, y_+]\}), \end{aligned} \tag{29}$$

where the outer expected value in the last expression is with respect to $f_0(\boldsymbol{X}, y_0, \boldsymbol{y})$ which, in turn, depends on true parameter vector $\boldsymbol{\theta}_0$. Similarly, the corresponding information matrix, equal to minus the second derivative matrix of $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$, is based on a single component of the sum in (20). More precisely, this information matrix is equal to $\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$, where

$$\begin{aligned}\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) &= -\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)] \\ &= E_0\{\boldsymbol{A}^*(\boldsymbol{X}) V^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_0,\boldsymbol{y})|\boldsymbol{X}, y_0, y_+]\boldsymbol{A}^*(\boldsymbol{X})'\},\end{aligned} \tag{30}$$

with the outer expected value in the last expression being with respect to $f_0(\boldsymbol{X}, y_0, y_+)$. This implies that $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$ is always concave and is strictly concave when $\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$ is of full rank. In this case, we denote by $\boldsymbol{\theta}_*$ the unique maximum of this function, which is obtained as the solution of $\boldsymbol{v}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*) = \boldsymbol{0}$. Moreover, also considering (26) and that $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$ is a concave function, the following theorem holds. This theorem directly derives from Theorem 2.7 of Newey and McFadden (1994); for related results, see also Akaike (1973) and White (1982).

Theorem 2. *Provided that the matrix $\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$ in (30) is of full rank and as $n \to \infty$, the pseudo conditional estimator $\hat{\boldsymbol{\theta}}$ exists with probability approaching 1 and $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_*$, where $\boldsymbol{\theta}_*$ is the unique maximum of $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$.*

A first important point is how to check the regularity condition that $\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$ is of full rank. This regularity condition may be empirically checked on the basis of the rank of $\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ computed to maximize $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$.

Another fundamental point is how to characterize the *pseudo true parameter vector* $\boldsymbol{\theta}_*$. We can easily realize that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ when $\gamma_0 = 0$ since, in this case, the data are generated under the static logit model which is a particular case of the proposed approximating model; see Eq. (11). This implies that $\boldsymbol{v}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$ is equal to $\boldsymbol{0}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and then the unique maximum of $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$ is at the true parameter vector $\boldsymbol{\theta}_0$ which, therefore, is correctly identified. Therefore, $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$ when $\gamma_0 = 0$.

On the other hand, when $\gamma_0 \neq 0$, $\hat{\boldsymbol{\theta}}$ converges to a point $\boldsymbol{\theta}_*$, the distance of which from $\boldsymbol{\theta}_0$ decreases as the distance of γ_0 from 0 decreases. If we knew the generating model, the point $\boldsymbol{\theta}_*$ could be found by a maximization algorithm of the type used to find the estimator $\hat{\boldsymbol{\theta}}$ and then we may obtain the asymptotic bias as $\boldsymbol{\theta}_* - \boldsymbol{\theta}_0$; see Section 5.2. However, in empirical applications, in which we have limited information on the generating model, we can perform a sensitivity analysis to figure out the maximum level of bias that we can expect. More detail on the computation of this asymptotic bias are given in Section 5.3 for the case of $T = 2$.

5.2. Asymptotic bias

Suppose that the generating model, and then the distribution $f_0(\alpha, \boldsymbol{X}, y_0)$ and the true parameter vector $\boldsymbol{\theta}_0$, is known. Then, we can compute by numerical integration, or by a Monte Carlo method, the expected log-likelihood function $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$ defined in (27) and maximize this function by performing a series of Newton–Raphson steps of type (21). Starting from $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, at the h-th of these steps, we update the previous solution, $\boldsymbol{\theta}^{(h-1)}$, by adding $\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)^{-1}\boldsymbol{v}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$, where the vector $\boldsymbol{v}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$ is computed through (29) and $\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)$ through (30). Note that, when an explicit solution is not available for computing $\bar{\boldsymbol{\beta}}_*$, which is the point at convergence of the estimator $\tilde{\boldsymbol{\beta}}$, we can find it by a similar maximization algorithm as above, which is based on the expected score vector $\boldsymbol{v}(\bar{\boldsymbol{\beta}})$ and the Fisher information $\boldsymbol{F}(\bar{\boldsymbol{\beta}})$. In particular, $\boldsymbol{v}(\bar{\boldsymbol{\beta}})$ may be computed by an expression similar to (29) considering that under the static model $\gamma = 0$; accordingly, $\boldsymbol{F}(\bar{\boldsymbol{\beta}})$ may be computed by an expression similar to (30). An example on how to apply this method to find $\boldsymbol{\theta}_*$ when we know the generating model, and then to obtain the asymptotic bias $\boldsymbol{\theta}_* - \boldsymbol{\theta}_0$, is provided in the following section.

In real applications, we observe sample values of the covariates and of the initial observation, that is $\boldsymbol{X}_i$ and y_{i0}, $i = 1, \ldots, n$. However, we have no information on the generating model, in particular concerning the distribution of the individual effects α_i. Then, in order to quantify the maximum expected bias of the proposed estimator we propose to perform a sensitivity analysis in which different distributions of these individual effects are considered and for each of these distributions we compute $\boldsymbol{\theta}_*$ and the corresponding distance from $\boldsymbol{\theta}_0$. For instance, we can use a normal distribution for these effects, with mean and variance chosen on a suitable grid of possible values.[4] Then, for each assumed distribution of the α_i, we perform an algorithm of the type described above to maximize an estimate of $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)]$, which is computed on the basis of the observed $\boldsymbol{X}_i$ and y_{i0}. This estimate is computed as

$$\hat{E}_0[\ell_i^*(\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}})] = \frac{1}{n}\sum_i E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}_*)|\boldsymbol{X}_i, y_{i0}],$$

where the expected value at rhs is with respect to the distribution of α_i and $p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0})$, assuming $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$ obtained on the observed sample. This function may be maximized by a Newton–Raphson algorithm similar to the one described above for the case in which we knew the true generating model. Starting from $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, this algorithm is based on steps of type $\hat{\boldsymbol{F}}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})^{-1}\hat{\boldsymbol{v}}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$, where

$$\begin{aligned}\hat{\boldsymbol{v}}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) &= \frac{1}{n}\sum_i \boldsymbol{A}^*(\boldsymbol{X}_i) E_0\{\boldsymbol{u}^*(y_0,\boldsymbol{y}) \\ &\quad - E^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_0,\boldsymbol{y})|\boldsymbol{X}_i, y_{i0}, y_+]|\boldsymbol{X}_i, y_{i0}\},\\ \hat{\boldsymbol{F}}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) &= \frac{1}{n}\sum_i \boldsymbol{A}^*(\boldsymbol{X}_i) \\ &\quad \times E_0\{V^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}[\boldsymbol{u}^*(y_0,\boldsymbol{y})|\boldsymbol{X}_i, y_{i0}, y_{i+}]|\boldsymbol{X}_i, y_{i0}\}\boldsymbol{A}^*(\boldsymbol{X}_i)'.\end{aligned}$$

In performing this sensitivity analysis, different values of the $\boldsymbol{\theta}_0$ around the estimate $\hat{\boldsymbol{\theta}}$ may also be tried, together with different formulations of the distribution of α_i.

5.3. Case of $T = 2$ time occasions

In order to illustrate the results in Sections 5.1 and 5.2, consider again the case of $T = 2$ time occasions dealt with in Section 4.2. In this case, the first derivative vector and the information matrix for $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\beta}_*)]$ have the same expressions as in (22) and (23), respectively, with each frequency $n_{y_0y_1y_2}$ and m_{y_0} substituted by the corresponding probabilities under the true model, denoted by $\pi_{y_0y_1y_2}$ and λ_{y_0}. In particular, it is important to note that

$$|\boldsymbol{F}^*(\boldsymbol{\theta}|\bar{\beta}_*)| = \frac{\exp[2\beta + (1-2\bar{q}_{*2})\gamma]\lambda_0\lambda_1}{k(0)^2 k(1)^2},$$

which is strictly positive if $\lambda_0 > 0$ and $\lambda_1 > 0$, implying that this matrix is of full rank and then Theorem 2 holds.

In this case it is easy to show that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ when $\gamma_0 = 0$ and then there is no state dependence. In fact, if $\gamma_0 = 0$ the first derivative vector of $E_0[\ell_i^*(\boldsymbol{\theta}|\bar{\beta}_*)]$ is equal to $\boldsymbol{0}$ when

$$\beta = \beta_0 = \log\frac{\pi_{001} + \pi_{101}}{\pi_{010} + \pi_{110}} = \mathrm{logit}\frac{\pi_{001} + \pi_{101}}{\lambda_0 + \lambda_1} \quad \text{and} \quad \gamma = 0.$$

[4] A referee suggested to choose a normal distribution for the individual effects, where the mean depends on the covariates. This is for allowing correlation between the regressors and the unobserved effects. However, we expect that the most challenging case in estimating a panel data model, such as a dynamic logit model, is when the individual effects are independent of the covariates, since otherwise part of the unobserved information is represented by these covariates.

$$\boldsymbol{v}^*(\boldsymbol{\theta}|\bar{\beta}_*) = \begin{pmatrix} \pi_{001}+\pi_{101}-\dfrac{\lambda_0(\pi_{001}+\pi_{101})}{\lambda_0+\lambda_1}-\dfrac{\lambda_1(\pi_{001}+\pi_{101})}{\lambda_0+\lambda_1} \\ \pi_{110}-\bar{q}_{*2}(\pi_{010}+\pi_{110})+\dfrac{\lambda_0\bar{q}_{*2}(\pi_{010}+\pi_{110})}{\lambda_0+\lambda_1}-\dfrac{\lambda_1(1-\bar{q}_{*2})(\pi_{010}+\pi_{110})}{\lambda_0+\lambda_1} \end{pmatrix} = \begin{pmatrix} 0 \\ \pi_{110}-\dfrac{\lambda_1(\pi_{010}+\pi_{110})}{\lambda_0+\lambda_1} \end{pmatrix}.$$

Box II.

Table 1
Asymptotic bias of $\hat{\beta}$ and $\hat{\gamma}$ under different generating models.

μ	σ^2	Estimator	True value of γ								
			−2.0	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5	2.0
−1	0.5	$\hat{\beta}$	0.568	0.362	0.196	0.074	0.000	−0.030	−0.024	0.004	0.038
		$\hat{\gamma}$	0.045	0.030	0.015	0.004	0.000	0.005	0.019	0.041	0.067
	1	$\hat{\beta}$	0.476	0.302	0.164	0.063	0.000	−0.029	−0.030	−0.014	0.006
		$\hat{\gamma}$	0.084	0.053	0.026	0.007	0.000	0.007	0.030	0.065	0.110
	2	$\hat{\beta}$	0.361	0.231	0.127	0.051	0.000	−0.028	−0.038	−0.036	−0.032
		$\hat{\gamma}$	0.139	0.084	0.039	0.010	0.000	0.011	0.042	0.093	0.160
0	0.5	$\hat{\beta}$	0.157	0.052	−0.006	−0.022	0.000	0.049	0.112	0.172	0.218
		$\hat{\gamma}$	0.067	0.041	0.019	0.005	0.000	0.004	0.015	0.030	0.045
	1	$\hat{\beta}$	0.131	0.047	−0.001	−0.015	0.000	0.037	0.084	0.131	0.164
		$\hat{\gamma}$	0.110	0.065	0.030	0.007	0.000	0.007	0.026	0.053	0.084
	2	$\hat{\beta}$	0.100	0.040	0.004	−0.008	0.000	0.022	0.051	0.077	0.093
		$\hat{\gamma}$	0.160	0.093	0.042	0.011	0.000	0.010	0.039	0.084	0.139
1	0.5	$\hat{\beta}$	−0.231	−0.225	−0.178	−0.100	0.000	0.107	0.207	0.287	0.339
		$\hat{\gamma}$	0.067	0.037	0.015	0.003	0.000	0.002	0.008	0.014	0.020
	1	$\hat{\beta}$	−0.192	−0.186	−0.147	−0.083	0.000	0.090	0.175	0.244	0.288
		$\hat{\gamma}$	0.110	0.061	0.026	0.006	0.000	0.005	0.017	0.033	0.049
	2	$\hat{\beta}$	−0.144	−0.137	−0.108	−0.060	0.000	0.066	0.128	0.176	0.203
		$\hat{\gamma}$	0.160	0.090	0.039	0.010	0.000	0.009	0.032	0.065	0.105

By substituting this solution in (29) we have the equation given in Box II. In particular, the second element is equal to 0 when $\gamma_0 = 0$ because in this case $\pi_{y_0y_1y_2}$ may be decomposed as the product between λ_{y_0} and $\pi_{y_1y_2}$. This example shows that, even with only $T = 2$ occasions, the method is able to identify the parameters of the dynamic logit model and consistently estimate these parameters when $\gamma_0 = 0$.

When $\gamma_0 \neq 0$ we can easily apply the method illustrated in Section 5.2 to find the asymptotic bias of the proposed estimator. In particular, suppose that the initial observation has probability 0.5 to be equal to 0 or to 1; moreover, we assume that $\alpha_i \sim N(\mu, \sigma^2)$ and we compute $\boldsymbol{\theta}_*$ for different values of μ, σ^2, and γ_0, with $\beta_0 = 1$ in all cases. The results are reported in Table 1 in terms of asymptotic bias.

5.4. Asymptotic distribution

Regularity conditions for asymptotic normality of the pseudo conditional estimator $\hat{\boldsymbol{\theta}}$ may be formulated by applying again the GMM theory; see, in particular, Newey and McFadden (1994, Sec. 6.1). The following Theorem results, where $\xrightarrow{d}$ stands for convergence in distribution and $V_0(\cdot)$ stands for variance under the true model.

Theorem 3. *Provided the condition in Theorem 2 holds, we have that*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V}^*(\boldsymbol{\theta}_*|\bar{\boldsymbol{\beta}}_*))$$

as $n \to \infty$, where $\boldsymbol{V}^*(\boldsymbol{\theta}_*|\bar{\boldsymbol{\beta}}_*)$ is the lower right submatrix of

$$E_0[\boldsymbol{H}_i(\bar{\boldsymbol{\beta}}_*, \boldsymbol{\theta}_*)]^{-1}V_0[\boldsymbol{g}_i(\bar{\boldsymbol{\beta}}_*, \boldsymbol{\theta}_*)]\{E_0[\boldsymbol{H}_i(\bar{\boldsymbol{\beta}}_*, \boldsymbol{\theta}_*)]^{-1}\}'.$$

Note that the lower right submatrix of $E_0[\boldsymbol{H}_i(\bar{\boldsymbol{\beta}}_*, \boldsymbol{\theta}_*)]$ is equal to the matrix $\boldsymbol{F}^*(\boldsymbol{\theta}_*|\bar{\boldsymbol{\beta}}_*)$ considered above. Moreover, since $\boldsymbol{H}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})/n$ converges in probability to $E_0[\boldsymbol{H}_i(\bar{\boldsymbol{\beta}}_*, \boldsymbol{\theta}_*)]$ and $\boldsymbol{S}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})/n$ converges in probability to $V_0[\boldsymbol{g}_i(\bar{\boldsymbol{\beta}}_*, \boldsymbol{\theta}_*)]$, with $\boldsymbol{g}_i(\bar{\boldsymbol{\beta}}_*, \boldsymbol{\theta}_*)$ defined in a similar way as in Section 4.3, the above theorem justifies the validity of the procedure based on (24) to obtain the standard errors for $\hat{\boldsymbol{\theta}}$.

6. Simulation study of the proposed estimator

In this section, we illustrate a simulation study carried out to assess the finite sample properties of the proposed pseudo estimator under the dynamic logit model in (2). In order to facilitate the comparison of our approach with alternative approaches, we follow the same simulation design adopted by Honoré and Kyriazidou (2000), to which we refer for a more detailed description of this design. The results also concern the confidence intervals that may be constructed around this estimator, as described in Section 4.3.

6.1. Simulation results

Similarly to Honoré and Kyriazidou (2000) we first consider a benchmark design and then some extended designs. Under the *benchmark design*, each sample is generated from a logit model with only one covariate which is based on the assumption

$$y_{it} = 1\{\alpha_i + x_{it}\beta + y_{i,t-1}\gamma + \varepsilon_{it} > 0\},\quad i = 1,\ldots,n,\ t = 1,\ldots,T,$$

with the initial condition

$$y_{i0} = 1\{\alpha_i + x_{it}\beta + \varepsilon_{it} > 0\},\quad i = 1,\ldots,n,$$

Table 2
Performance of the pseudo conditional estimator under some benchmark simulation designs with $T = 3$ and $\beta = 1$. Percentages are referred to the ratio between the actual sample size and the nominal one.

γ	n	Estimation of β				Estimation of γ			
		Mean bias	RMSE	Median bias	MAE	Mean bias	RMSE	Median bias	MAE
0.25	250	0.025	0.142	0.009	0.088	−0.012	0.389	−0.017	0.253
(60%)	500	0.009	0.093	0.002	0.062	−0.011	0.282	−0.010	0.195
	1000	0.004	0.066	0.001	0.043	0.000	0.195	0.005	0.127
0.50	250	0.025	0.146	0.007	0.087	−0.016	0.394	−0.027	0.257
(57%)	500	0.006	0.094	0.000	0.060	−0.010	0.287	−0.004	0.190
	1000	0.002	0.066	−0.002	0.046	−0.003	0.198	−0.001	0.132
1.00	250	0.025	0.151	0.009	0.094	−0.013	0.429	−0.027	0.278
(52%)	500	0.006	0.099	0.001	0.064	−0.017	0.306	−0.023	0.204
	1000	−0.001	0.068	−0.003	0.047	−0.009	0.210	−0.010	0.142
2.00	250	0.037	0.181	0.015	0.107	0.022	0.564	−0.022	0.365
(42%)	500	0.011	0.116	−0.002	0.073	−0.014	0.368	−0.024	0.249
	1000	−0.005	0.077	−0.012	0.050	−0.028	0.256	−0.033	0.173

Table 3
Performance of the pseudo conditional estimator under some benchmark simulation designs with $T = 7$ and $\beta = 1$. Percentages are referred to the ratio between the actual sample size and the nominal one.

γ	n	Estimation of β				Estimation of γ			
		Mean bias	RMSE	Median bias	MAE	Mean bias	RMSE	Median bias	MAE
0.25	250	0.007	0.060	0.004	0.042	0.002	0.157	0.004	0.102
(92%)	500	0.004	0.042	0.002	0.028	0.004	0.117	0.004	0.078
	1000	0.000	0.030	0.001	0.021	0.000	0.081	0.000	0.055
0.50	250	0.007	0.062	0.005	0.043	0.006	0.160	−0.002	0.106
(91%)	500	0.004	0.042	0.004	0.029	0.006	0.117	0.004	0.078
	1000	0.000	0.030	0.000	0.021	0.001	0.084	0.001	0.056
1.00	250	0.009	0.064	0.004	0.042	0.013	0.171	0.003	0.111
(87%)	500	0.005	0.045	0.004	0.030	0.009	0.122	0.004	0.082
	1000	−0.001	0.031	−0.001	0.021	0.002	0.088	0.000	0.061
2.00	250	0.010	0.073	0.008	0.050	0.017	0.200	0.009	0.134
(76%)	500	0.008	0.051	0.006	0.034	0.010	0.148	0.005	0.101
	1000	0.002	0.035	0.001	0.024	0.004	0.104	0.000	0.069

where $T = 3$, $\beta = 1$, and $\gamma = 0.5$. Each covariate x_{it} is drawn from a Normal distribution with mean 0 and variance $\pi^2/3$, whereas each α_i is generated as $\sum_{t=0}^{3} x_{it}/4$.

To study the sensitivity of the results on the simulation design, we also consider a number of time occasions T equal to 7 and different values of γ (0.25, 0.5, 1, 2). About the choice of γ, consider that in typical microeconomic analyses via dynamic logit or probit models, which are available in the literature, values of the state dependence parameters result positive and not higher than 2 (on the logit scale). For instance, Hyslop (1999), in analyzing data about female participation to the labor market, found that a reliable estimate of the state dependence parameter is close to 1 on the probit scale, which corresponds to a value around 1.6 on the logit scale. Another example is represented by the application described by Hsiao (2005, Sec. 7.5) about brand choice by a sample of customers. In this case, reliable estimates of the state dependence parameter are around 1.2.

Following Honoré and Kyriazidou (2000), we also assume different distributions for the covariate. In particular we consider four further designs. In the first one, we generate each x_{it} from a $\chi^2(1)$ distribution transformed so to have mean 0 and variance $\pi^2/3$. In the second design, the model is estimated with three more covariates generated from the same Normal distribution adopted to generate x_{it}. In the third and fourth designs, the covariate is generated as $x_{it} = \rho(\xi + 0.1t + \zeta_{it})$, with ρ and ξ suitably chosen and where $\zeta_{i0}, \ldots, \zeta_{iT}$ follow a Gaussian AR(1) process with autoregressive coefficient equal to 0.5, normalized so to have variance $\pi^2/3$, with $T = 3$ and $T = 7$. Finally, following the suggestion of one of the referees, we also try different ways to generate the incidental parameters. In particular, we assume $\alpha_i = \mu + \sigma \sum_{t=0}^{3} x_{it}/4$ for $i = 1, \ldots, n$, with $\mu = 0, 1, 2$ and $\sigma^2 = 0.5, 1, 1.5, 2$.

For each model described above, we simulated 1000 samples of size n, with $n = 250, 500, 1000$. On the basis of each sample we estimated the structural parameters of the logit model by the proposed pseudo conditional estimator $\hat{\boldsymbol{\theta}}$. For these parameters we also constructed 80% and 95% confidence intervals as described in Section 4.3. The results in terms of *mean bias*, *root mean squared error* (RMSE), *median bias*, and *median absolute error* (MAE) of the estimators are shown in Tables 2 and 3 for the benchmark design and in Tables 5 and 7 for the other designs. For each value of γ, these tables also show the ratio between the actual sample size and the nominal sample size n.[5] The results, in terms of actual coverage level of the confidence intervals, are displayed in Table 4 for the benchmark design and in Table 6 for some of the other designs.

As for the bias of the pseudo conditional estimator $\hat{\beta}$, Tables 2 and 3 show that this bias is always negligible. Moreover, about its efficiency, we note that both RMSE and MAE decrease with n at a rate close to $\sqrt{n}$ and much faster with T. Both RMSE and MAE moderately increase with γ. One of the main reasons of this is that the actual sample size tends to increase with T and decreases with γ when γ is positive. The picture for the pseudo conditional estimator $\hat{\gamma}$ is quite similar. Its bias is very close to 0; moreover, both RMSE and MAE of $\hat{\gamma}$ moderately increase with γ, decrease as n grows at a rate close to $\sqrt{n}$ and much faster with T.

The good performance of the pseudo estimator is confirmed by the behavior of the confidence intervals. In particular, as shown in Table 4, the actual coverage level of the confidence intervals for β is always very close to the nominal one. Similar conclusions may be drawn regarding the confidence intervals for γ.

[5] This ratio is computed as the expected proportion of response configurations $\boldsymbol{y}_i$ such that $0 < y_{i+} < T$.

Table 4
Coverage levels of the confidence intervals based on the pseudo conditional estimator under some benchmark simulation designs, with $\beta = 1$.

γ	n	$T = 3$				$T = 7$			
		Interval for β		Interval for γ		Interval for β		Interval for γ	
		80%	95%	80%	95%	80%	95%	80%	95%
0.25	250	0.81	0.96	0.80	0.95	0.80	0.95	0.79	0.94
	500	0.81	0.96	0.79	0.95	0.79	0.94	0.80	0.95
	1000	0.80	0.95	0.80	0.95	0.81	0.95	0.81	0.96
0.50	250	0.82	0.95	0.81	0.95	0.79	0.96	0.80	0.95
	500	0.82	0.95	0.80	0.96	0.80	0.95	0.81	0.95
	1000	0.80	0.95	0.80	0.95	0.79	0.95	0.80	0.96
1.00	250	0.82	0.95	0.80	0.95	0.80	0.95	0.81	0.94
	500	0.80	0.95	0.80	0.95	0.79	0.96	0.80	0.95
	1000	0.79	0.97	0.81	0.95	0.80	0.95	0.80	0.95
2.00	250	0.82	0.96	0.81	0.95	0.81	0.95	0.81	0.95
	500	0.82	0.95	0.81	0.95	0.79	0.95	0.81	0.95
	1000	0.79	0.95	0.81	0.95	0.81	0.94	0.80	0.94

From Table 5 we observe that, under the simulation designs based on different distributions for the covariates, the pseudo conditional estimator has essentially the same behavior it has under the benchmark design. Even when the estimator performs worse, in terms of bias and/or efficiency, with respect to the benchmark design, differences are small. This happens for the $\chi^2(1)$ design (limited to the efficiency of $\hat{\beta}$) and for the additional regressors design. Occasionally, the proposed estimator also performs better than under the benchmark design. Limited to $\hat{\gamma}$, this happens, for instance, under the $\chi^2(1)$ design.

For what concerns the confidence intervals, we observe from Table 6 that, even under the simulation designs based on alternative distributions for the covariates, the actual coverage level is always very close to the nominal level for both parameters β and γ. This confirms the quality of the method proposed in Section 4.3 to obtain standard errors and to construct confidence intervals, already noticed for the benchmark design.

Finally, on the basis of the results in Table 7, we conclude that the bias and the efficiency of $\hat{\beta}$ and $\hat{\gamma}$ slightly worsen as the mean of the α_i parameters rises, but they are rather insensible to changes of the variance.

6.2. Comparison with alternative estimators

An important issue is how the proposed pseudo conditional estimator performs in comparison to the weighted conditional estimator of Honoré and Kyriazidou (2000) and the bias corrected estimator of Carro (2007). We then compare the simulation results obtained by these authors with those illustrated above. We also present the results for the *infeasible logit estimator* that uses the fixed effect as one of the explanatory variables; see Honoré and Kyriazidou (2000) for details. This comparison is summarized in Table 8, which, for certain reference situations and for both β and γ, shows the median bias and the MAE of our estimator in comparison to those of the other estimators. For all these estimators, the table also shows the rate between the actual sample size and the nominal sample size.[6]

An advantage of our estimator over the alternative estimators, in terms of bias and efficiency, clearly emerges from the results in Table 8. In particular, with respect to the weighted conditional estimator of Honoré and Kyriazidou (2000), our estimator $\hat{\beta}$ of β always has a smaller median bias and MAE. Moreover, especially from the point of view of the efficiency, the advantage of our estimator increases with γ and n and decreases with T. For instance, with $\gamma = 2$, $T = 3$, and $n = 1000$, $\hat{\beta}$ has median bias equal to -0.012 and MAE equal to 0.050; the weighted conditional estimator, instead, has median bias equal to 0.113 and MAE equal to 0.136. A similar advantage may be observed in estimating γ. Even in this case we observe that our estimator $\hat{\gamma}$ has always smaller median bias and MAE than the conditional weighted estimator. This advantage clearly increases with γ, whereas there is not a clear trend in T and n.

The main explanation that we can give for the above results is that the actual sample size exploited in our approach is always much larger than that exploited in the approach of Honoré and Kyriazidou (2000). This difference increases with γ and T. For instance, with $\gamma = 0.5$ and $T = 3$, the actual sample size used in our approach is about 1.5 times that used in their approach. This ratio increases to about 2.1 for $\gamma = 0.5$ and $T = 7$ and to 2.2 for $\gamma = 2$ and $T = 7$. Note however that the gain in median bias and MAE does not closely follows the gain in the actual sample size. Therefore, other factors have to be taken into consideration which affect the performance of the two estimators in a way that depends on both γ and T. We recall, in particular, that the performance of our estimator depends on the quality of the approximation we are relying on, whereas the performance of the estimator of Honoré and Kyriazidou (2000) also depends on the fact that the response configurations are differently weighted on the basis of the corresponding covariate configurations and that, for $T > 3$, they are indeed relying on a pairwise likelihood.

In comparison to the bias corrected estimator of Carro (2007), our estimator $\hat{\beta}$ always has a smaller median bias, but not always a smaller MAE. At least in terms of efficiency, the relative performance of the two estimators seems to be rather insensitive to γ; moreover, the advantage of our estimator increases with n, but it has not a clear trend in T. The situation is different when the parameter of interest is γ. In this case our estimator $\hat{\gamma}$ always outperforms the estimator of Carro (2007) in terms of bias and efficiency. The advantage of the proposed approach increases with γ and n and decreases with T and in certain cases is evident. For instance, with $\gamma = 2$, $T = 3$, and $n = 1000$, our estimator $\hat{\gamma}$ has median bias of -0.033 and MAE equal to 0.173, whereas the alternative estimator has median bias equal to -1.265 and MAE equal to 1.252. In order to explain this advantage, we recall that the approach of Carro (2007) ensures a reduced bias for long panels, whereas for short panels there may be a strong bias, especially in estimating γ. In any case, his approach exploits the same actual sample size as ours.

Finally, the conditional estimator performs quite well in relation to the infeasible estimator, being better when n is bigger and T is larger.

7. Extensions

In this Section, we extend the pseudo conditional approach to two more general cases. The first case concerns the inclusion, in the logit model, of more than one lagged response variable among the regressors, so as to extend the first-order Markovian assumption. The second case concerns categorical response variables having more than two categories.

7.1. Inclusion of more lagged response variables

The first-order Markovian assumption for the response variables is here relaxed to allow for longer dynamics. In particular, we illustrate the case of two lags.

[6] For the weighted conditional estimator, this rate is computed as the expected proportion of pairs of response variables (y_{is}, y_{it}), $0 < s < t < T$, such that $y_{is} + y_{it} = 1$.

Table 5
Performance of the pseudo conditional estimator under different simulation designs, with $\beta = 1$ and $\gamma = 0.5$. Percentages are referred to the ratio between the actual sample size and the nominal one.

Type of design	n	Estimation of β				Estimation of γ			
		Mean bias	RMSE	Median bias	MAE	Mean bias	RMSE	Median bias	MAE
$\chi^2(1)$	250	0.017	0.160	0.001	0.104	−0.015	0.336	−0.033	0.224
Regressors	500	0.010	0.108	0.001	0.072	−0.013	0.229	−0.016	0.156
($T = 3$, 56%)	1000	0.002	0.077	−0.003	0.050	−0.007	0.171	−0.012	0.120
Additional	250	0.050	0.154	0.038	0.094	−0.008	0.421	−0.023	0.290
Regressors	500	0.015	0.096	0.010	0.060	0.000	0.275	−0.008	0.183
($T = 3$, 57%)	1000	0.010	0.063	0.010	0.043	−0.016	0.191	−0.016	0.127
Trending	250	0.030	0.170	0.016	0.104	−0.030	0.440	−0.035	0.286
Regressors	500	0.013	0.116	0.002	0.077	−0.030	0.293	−0.025	0.200
($T = 3$, 42%)	1000	0.001	0.080	−0.004	0.054	−0.014	0.207	−0.011	0.137
Trending	250	0.006	0.072	0.004	0.050	−0.002	0.180	−0.003	0.117
Regressors	500	0.002	0.049	0.002	0.034	0.000	0.124	0.002	0.078
($T = 7$, 78%)	1000	0.000	0.036	−0.001	0.023	−0.001	0.090	−0.004	0.059

Table 6
Coverage levels of the confidence intervals based on the pseudo conditional estimator under different simulation designs, with $\beta = 1$ and $\gamma = 0.5$.

Type of design	n	Interval for β		Interval for γ	
		80%	95%	80%	95%
$\chi^2(1)$	250	0.81	0.95	0.80	0.95
Regressors	500	0.82	0.95	0.79	0.95
	1000	0.83	0.95	0.80	0.95
Additional	250	0.82	0.95	0.80	0.94
Regressors	500	0.80	0.95	0.80	0.95
	1000	0.80	0.95	0.79	0.95
Trending	250	0.82	0.96	0.80	0.95
Regressors	500	0.81	0.95	0.81	0.95
($T = 3$)	1000	0.79	0.95	0.80	0.94
Trending	250	0.80	0.96	0.80	0.94
Regressors	500	0.81	0.95	0.81	0.94
($T = 7$)	1000	0.79	0.94	0.79	0.95

Including two lags, the dynamic logit model described in Section 2.1 becomes

$$p(y_{it}|\alpha_i, \boldsymbol{X}_i, y_{i,-1}, \dots, y_{i,t-1}) = p(y_{it}|\alpha_i, \boldsymbol{x}_{it}, y_{i,t-2}, y_{i,t-1}) \\ = \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma_1 + y_{i,t-2}\gamma_2)]}{1+\exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma_1 + y_{i,t-2}\gamma_2)}, \\ i = 1, \dots, n, \; t = 1, \dots, T,$$

with γ_1 and γ_2 having an obvious interpretation, and $y_{i,-1}$ and y_{i0} denoting the two initial observations, assumed to be exogenous. Under this assumption, it is straightforward to write the distribution of $\boldsymbol{y}_i$, given α_i, $\boldsymbol{X}_i$, $y_{i,-1}$, and y_{i0}, as

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i,-1}, y_{i0}) \\ = \frac{\exp\left(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i*1}\gamma_1 + y_{i*2}\gamma_2\right)}{\prod_t [1+\exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma_1 + y_{i,t-2}\gamma_2)]},$$

where $y_{i*1} = \sum_t y_{i,t-1}y_{it}$ and $y_{i*2} = \sum_t y_{i,t-2}y_{it}$.

The logarithm of the denominator of the joint probability above may be approximated by a first-order Taylor-series expansion around $\alpha_i = \bar{\alpha}_i$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, and $\gamma_1 = \gamma_2 = 0$, obtaining

$$\sum_t \log[1+\exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma_1 + y_{i,t-2}\gamma_2)] \\ \approx \sum_t \{\log[1+\exp(\bar{\alpha}_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}})] + \bar{q}_{it}[\alpha_i - \bar{\alpha}_i + \boldsymbol{x}_{it}'(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})]\} \\ + \sum_t \bar{q}_{it}(y_{i,t-1}\gamma_1 + y_{i,t-2}\gamma_2),$$

with $\bar{q}_{it}$ defined as in (9). Therefore, after some algebra, we find that $p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i,-1}, y_{i0})$ may be approximated by the equation which is given in Box III, with z_{i*1} and z_{i*2} defined in the usual way. The approximating model is therefore a quadratic exponential model in which the main effect parameter for y_{it} is equal to $\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} - (\gamma_1+\gamma_2)\sum_t \bar{q}_{it}$ when $t = 1, \dots, T-2$, to $\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} - \gamma_1 \sum_t \bar{q}_{it}$ when $t = T-1$, and to $\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}$ when $t = T$; moreover, the two-way interaction effect for (y_{is}, y_{it}) is equal to γ_1 when $t = s+1$, to γ_2 when $t = s+2$, and to 0 otherwise.

The main advantage of the approximating model is again that of having a minimal sufficient statistic for the individual parameters α_i. These sufficient statistics are y_{i+}, $i = 1, \dots, n$, so that the conditional distribution of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$, $y_{i,-1}$, y_{i0}, and y_{i+} does not depend on α_i. Consequently, the structural parameters may be estimated by maximizing the pseudo likelihood based on this conditional distribution in a way similar to that outlined in Section 4.1. The resulting pseudo conditional estimator has essentially the same asymptotic properties of the initial pseudo conditional likelihood estimator; see Section 5.

7.2. *Dealing with response variables having more categories*

When the response variables have more than two categories, and these categories are ordered, the formulation in Section 2.1 may be naturally extended by using in (1) a different definition for the indicator function. More precisely, let J denote the number of response categories, from 1 to J. Then, the indicator function to be used is such that it yields j when its argument, in our case $\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma + \varepsilon_{it}$, is between two cutpoints c_{j-1} and c_j, with $c_0 = -\infty$ and $c_J = \infty$; more sophisticated ways may be also adopted to include state dependence. When the errors terms ε_{it} are assumed to have a logistic distribution, a model that may be referred to as the *dynamic ordered logit model* results. Under this model, we have an expression of type (2) for the joint probability of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$ and y_{i0} which, however, is based on cumulative (or global) logits. For the definition of logits of this type see McCullagh (1980) and Agresti (2002).

To apply the proposed approach to estimate the above model it is convenient to follow a general idea found in the literature (Mukherjee et al., 2008; Baetschmann et al., 2011), which consists of collapsing the response categories in different ways, so as to make the response variables binary. In particular, for $j = 1, \dots, J-1$, we can transform each y_{it} in the binary response variable $y_{it}^{(j)}$ defined as follows

$$y_{it}^{(j)} = \begin{cases} 0 & \text{if } y_{it} \le j, \\ 1 & \text{if } y_{it} > j. \end{cases}$$

Table 7
Performance of the pseudo conditional estimator under different values of mean (μ) and the variance (σ^2) of α_i, with $T = 3$, $\beta = 1$, and $\gamma = 1$.

σ^2	μ	n	Estimation of β				Estimation of γ			
			Mean bias	RMSE	Median bias	MAE	Mean bias	RMSE	Median bias	MAE
0.5	0	250	0.024	0.144	0.006	0.088	−0.008	0.417	−0.019	0.271
		1000	−0.002	0.066	−0.005	0.046	−0.005	0.198	−0.010	0.133
	1	250	0.027	0.156	0.008	0.095	−0.033	0.449	−0.063	0.292
		1000	−0.002	0.068	−0.003	0.048	−0.034	0.216	−0.036	0.147
	2	250	0.044	0.191	0.024	0.119	−0.048	0.598	−0.075	0.374
		1000	−0.002	0.081	−0.007	0.055	−0.060	0.275	−0.063	0.193
1	0	250	0.025	0.151	0.009	0.094	−0.013	0.429	−0.027	0.278
		1000	−0.001	0.068	−0.003	0.047	−0.009	0.210	−0.010	0.142
	1	250	0.027	0.158	0.007	0.097	−0.029	0.470	−0.063	0.319
		1000	−0.003	0.070	−0.002	0.047	−0.033	0.225	−0.036	0.149
	2	250	0.046	0.199	0.022	0.118	−0.039	0.576	−0.055	0.373
		1000	0.001	0.082	−0.005	0.056	−0.056	0.276	−0.049	0.186
1.5	0	250	0.028	0.155	0.009	0.095	−0.024	0.452	−0.035	0.295
		1000	−0.002	0.070	−0.003	0.048	−0.012	0.216	−0.008	0.146
	1	250	0.034	0.168	0.011	0.103	−0.041	0.490	−0.059	0.321
		1000	−0.002	0.074	−0.004	0.047	−0.035	0.226	−0.038	0.145
	2	250	0.044	0.198	0.017	0.121	−0.057	0.587	−0.069	0.397
		1000	0.000	0.080	−0.003	0.050	−0.057	0.276	−0.059	0.176
2	0	250	0.032	0.160	0.012	0.099	−0.028	0.468	−0.036	0.306
		1000	−0.001	0.073	−0.002	0.049	−0.012	0.225	−0.015	0.153
	1	250	0.033	0.170	0.010	0.101	−0.038	0.504	−0.072	0.340
		1000	−0.001	0.078	−0.003	0.052	−0.035	0.235	−0.039	0.155
	2	250	0.042	0.199	0.016	0.118	−0.045	0.577	−0.071	0.400
		1000	0.001	0.084	−0.003	0.053	−0.055	0.274	−0.060	0.178

$$p^*(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i,-1}, y_{i0}) = \frac{\exp\left(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} - \sum_{t>1}\bar{q}_{it}y_{i,t-1}\gamma_1 - \sum_{t>2}\bar{q}_{it}y_{i,t-2}\gamma_2 + y_{i*1}\gamma_1 + y_{i*2}\gamma_2\right)}{\sum_{\boldsymbol{z}}\exp\left(z_+\alpha_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\beta} - \sum_{t>1}\bar{q}_{it}z_{t-1}\gamma_1 - \sum_{t>2}\bar{q}_{it}z_{t-2}\gamma_2 + z_{i*1}\gamma_1 + z_{i*2}\gamma_2\right)},$$

Box III.

Then, for each of these dichotomizations we obtain a pseudo conditional log-likelihood as in (15), denoted by $\ell^{*(j)}(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$, with $\bar{\boldsymbol{\beta}}$ suitably chosen. Then, we define an overall pseudo conditional log-likelihood function as

$$\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \sum_j \ell^{*(j)}(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}).$$

This function is maximized with respect to $\boldsymbol{\theta}$ by a simple Newton–Raphson algorithm. Note that $\boldsymbol{\theta}$ also includes the cutpoints $c_1, \ldots, c_{J-1}$, which are then parameters to estimate. Even if a deeper study is necessary, the resulting estimator is expected to have asymptotic properties similar those illustrated in Section 5.

Alternatively, the dynamic logit model may be extended by assuming

$$p(y_{it}|\alpha_i, \boldsymbol{X}_i, y_{i0}, \ldots, y_{i,t-1}) = p(y_{it}|\alpha_i, \boldsymbol{x}_{it}, y_{i,t-1})$$
$$= \frac{\exp(\alpha_{iy_{it}} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_{y_{it}} + \gamma_{y_{i,t-1}y_{it}})}{\sum_j \exp(\alpha_{ij} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_j + \gamma_{y_{i,t-1}j})},$$
$$i = 1, \ldots, n,\ t = 1, \ldots, T.$$

This is the *dynamic multinomial logit model*, which is based on the incidental parameters α_{ij}, $i = 1, \ldots, n, j = 1, \ldots, J$, and the structural parameters $\boldsymbol{\beta}_j$, $j = 1, \ldots, J$, and γ_{hj}, $h, j = 1, \ldots, J$. Suitable constraints, such as $\boldsymbol{\beta}_1 = \boldsymbol{0}$, are assumed on these parameters in order to ensure identifiability.

Under the above formulation, the method illustrated in Section 4 may be directly applied in order to derive an approximation of the joint probability of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$ and y_{i0}. In particular, it may be easily shown that the approximating model has again sufficient statistics for the incidental parameters α_{ij} which are $y_{i+}^{(j)} = \sum_t 1\{y_{it} = j\}$, $i = 1, \ldots, n, j = 1, \ldots, J$. In practice, $y_{i+}^{(j)}$ is equal to the number of response variables y_{it} which, during the period of observation, are equal to j. On the basis of the log-likelihood resulting from conditioning on these sufficient statistics, we obtain a pseudo conditional estimator of the structural parameters. We defer the study of the properties of this estimator to future research.

8. Conclusions

In this paper, we propose a pseudo conditional likelihood approach for a dynamic logit model which allows for unobserved heterogeneity and individual covariates. The proposed approach is based on approximating this model, which is referred to as the *true model*, by a version of the quadratic exponential model (Cox, 1972), which corresponds to the *approximating model*. On the basis of the latter we construct a pseudo conditional likelihood which does not depend on the incidental parameters for the unobserved heterogeneity. This is obtained by conditioning on simple sufficient statistics which are the sums of the response variables for every sample unit. The pseudo conditional estimator of the structural parameters, for the covariates and the state dependence, is obtained by maximizing the conditional log-likelihood of the approximating model, given the sufficient statistics, by means of a Newton–Raphson algorithm.

The main feature of the above estimator is that it is simpler to use and performs better than alternative estimators proposed in the literature. In particular, with respect to the weighted conditional estimator of Honoré and Kyriazidou (2000), which may be considered as a benchmark in this literature, our estimator:

Table 8
Comparison between the infeasible logit estimator (In), the estimator of Honoré and Kyriazidou (2000, HK), the estimator of Carro (2007, C), and the proposed pseudo conditional estimator (*P*). Percentages in the first two columns are referred to actual sample size under the last three approaches (the first to the HK estimator and the second to the *C* and *P* estimators).

γ	T	n	Estimator	Estimation of β		Estimation of γ	
				Median bias	MAE	Median bias	MAE
0.5	3	250	In	0.006	0.051	−0.005	0.103
			HK	0.076	0.154	−0.039	0.403
(37%–57%)			C	−0.054	0.068	−0.554	0.554
			P	0.007	0.087	−0.027	0.257
		1000	In	0.000	0.026	−0.002	0.051
			HK	0.038	0.086	−0.035	0.178
			C	−0.057	0.057	−0.563	0.563
			P	−0.002	0.046	−0.001	0.132
	7	250	In	0.002	0.033	−0.004	0.063
			HK	0.014	0.050	−0.053	0.131
(43%–91%)			C	0.012	0.039	−0.106	0.127
			P	0.005	0.043	−0.002	0.106
		1000	In	0.000	0.018	−0.001	0.031
			HK	0.009	0.027	−0.041	0.075
			C	0.015	0.022	−0.097	0.098
			P	0.000	0.021	0.001	0.056
2	3	250	In	0.006	0.057	0.003	0.120
			HK	0.196	0.251	−0.056	0.620
(26%–42%)			C	−0.086	0.079	−1.181	0.990
			P	0.015	0.107	−0.022	0.365
		1000	In	0.001	0.025	0.002	0.061
			HK	0.113	0.136	−0.148	0.321
			C	−0.061	0.060	−1.265	1.252
			P	−0.012	0.050	−0.033	0.173
	7	250	In	0.005	0.039	−0.003	0.075
			HK	0.016	0.064	−0.195	0.227
(34%–76%)			C	0.019	0.045	−0.226	0.227
			P	0.008	0.050	0.009	0.134
		1000	In	−0.001	0.018	−0.002	0.038
			HK	0.016	0.034	−0.160	0.164
			C	0.016	0.023	−0.218	0.218
			P	0.001	0.024	0.000	0.069

(i) does not require a kernel function for weighting the response configurations and (ii) it may be also used with at least two time occasions, instead of at least three, and in the presence of time dummies. A more important aspect is that our estimator usually has a smaller bias and a greater efficiency. This conclusion is based on a simulation study that we performed along the same lines as Honoré and Kyriazidou (2000). In particular, we notice that our estimator has a surprisingly low bias under each scenario considered in the simulation study. It also has a root mean square error and a median absolute error that decrease, as n grows, at a rate close to $\sqrt{n}$. Moreover, the advantage in terms of bias and efficiency over the estimator of Honoré and Kyriazidou (2000) is higher when there is a strong state dependence. An intuitive explanation of the better performance of our approach is that it is based on a conditional likelihood to which a larger number of response configurations contribute (actual sample size) with respect to the likelihood on which the estimator of Honoré and Kyriazidou (2000) is based. This conclusion does not contradict the result of Hahn (2001), who showed that, for the dynamic logit with time dummies and $T = 3$, there does not exist any $\sqrt{n}$-consistent estimator. In fact, our estimator, being based on a conditional likelihood of an approximating model, does not belong to the class of estimators consider by Hahn (2001), which are based on the conditional likelihood of the true model. An advantage, especially in the estimation of the state dependence parameter, is also observed in comparison to the bias corrected estimator proposed by Carro (2007), which for short panels may have a considerable bias.

In this paper, we also suggest how to obtain standard errors for the pseudo conditional likelihood estimator. These standard errors are estimated in a robust way by using a sandwich formula (White, 1982). On the basis of these standard errors we can construct confidence intervals for the structural parameters of the true model.

Finally, we outline how to extend the pseudo conditional likelihood approach to two more complex cases which involve longer dynamics and categorical response variables (ordinal and non-ordinal) having more than two categories. We think that it is also possible to extend the proposed approach to other models, such as the dynamic probit model, but we leave this extension to future research.

Appendix. Blocks of the second derivative matrix $\boldsymbol{H_i}(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta})$

We have that

$$\nabla_{\bar{\beta}\bar{\beta}}\ell_i(\bar{\boldsymbol{\beta}}) = \boldsymbol{A}(\boldsymbol{X}_i)V_{\bar{\beta}}[\boldsymbol{u}(\boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}]\boldsymbol{A}(\boldsymbol{X}_i)',$$

where $\boldsymbol{u}(\boldsymbol{y}_i) = (y_{i2}, \dots, y_{iT})'$, $\boldsymbol{A}(\boldsymbol{X}_i) = \boldsymbol{X}_i\boldsymbol{D}'$, and $V_{\bar{\beta}}(\cdot)$ denotes the variance under the static logit model. Moreover, we have that

$$\nabla_{\theta\theta}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \boldsymbol{A}^*(\boldsymbol{X}_i)V^*_{\theta|\bar{\beta}}[\boldsymbol{u}^*(y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}]\boldsymbol{A}^*(\boldsymbol{X}_i)'.$$

Finally, the block $\nabla_{\theta\bar{\beta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ is rather complicated to compute analytically. Therefore, we prefer to rely on a numerical derivative of the score of $\boldsymbol{\theta}$ with respect to $\bar{\boldsymbol{\beta}}$.

References

Agresti, A., 2002. Categorical Data Analysis, 2nd edition. John Wiley & Sons, New York.
Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Proceedings of the Second International Symposium of Information Theory. Akademiai Kiado, Budapest, pp. 267–281.
Andersen, E.B., 1970. Asymptotic properties of conditional maximum-likelihood estimators. Journal of Royal Statistical Society, B 32, 283–301.
Andersen, E.B., 1972. The numerical solution of a set of conditional estimation equations. Journal of the Royal Statistical Society, B 34, 42–54.
Baetschmann, G., Staub, K.E., Winkelmann, R., 2011. Consistent estimation of the fixed effects ordered logit model. IZA Discussion paper, 5443.
Bartolucci, F., 2010. On the conditional logistic estimator in two-arm experimental studies with non-compliance and before–after binary outcomes. Statistics in Medicine 29, 1411–1429.
Barndoff-Nielsen, O.E., 1978. Information and Exponential Families in Statistical Theory. John Wiley & Sons, New York.
Bartolucci, F., Farcomeni, A., 2009. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. Journal of the American Statistical Association 104, 816–831.
Bartolucci, F., Nigro, V., 2010. A dynamic model for binary panel data with unobserved heterogeneity admitting a root-n consistent conditional estimator. Econometrica 78, 719–733.
Bartolucci, F., Pennoni, F., 2007. On the approximation of the quadratic exponential distribution in a latent variable context (article). Biometrika 94, 745–754.
Carro, J., 2007. Estimating dynamic panel data discrete choice models with fixed effects. Journal of Econometrics 140, 503–528.
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.
Cox, D.R., 1972. The analysis of multivariate binary data. Applied Statistics 21, 113–120.
Cox, D.R., Wermuth, N., 1994. A note on the quadratic exponential binary distribution. Biometrika 81, 403–408.
Diggle, P.J., Heagerty, P., Liang, K.-Y., Zeger, S.L., 2002. Analysis of Longitudinal Data. Oxford University Press, New York.
Feller, W., 1943. On a General class of 'contagious' distributions. Annals of Mathematical Statistics 14, 389–400.
Fernandez-Val, I., 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit model. Journal of Econometrics 150, 71–85.
Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80, 27–38.
Hahn, J., 2001. The information bound of a dynamic panel logit model with fixed effects. Econometric Theory 17, 913–932.
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., 1981a. Statistical models for discrete panel data. In: Manski, C.F., McFadden, D.L. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, pp. 114–178.
Heckman, J.J., 1981b. Heterogeneity and state dependence. In: Rosen, S. (Ed.), Studies in Labor Markets. University of Chicago Press, Chicago, pp. 91–140.
Heckman, J.J., 1981c. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski, C.F., McFadden, D.L. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT press, Cambridge, MA, pp. 179–195.
Honoré, B.E., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68, 839–874.
Honoré, B.E., Tamer, E., 2006. Bounds on parameters in panel dynamic discrete choice models. Econometrica 74, 611–629.
Hsiao, C., 2005. Analysis of Panel Data, second ed. Cambridge University Press, New York.
Hyslop, D.R., 1999. State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. Econometrica 67, 1255–1294.
Magnac, T., 2004. Panel binary variables and sufficiency: generalizing conditional logit. Econometrica 72, 1859–1876.
McCullagh, P., 1980. Regression models for ordinal data (with discussion). Journal of the Royal Statistical Society. Series B 42, 109–142.
McCullagh, P., Tibshirani, R., 1990. A simple method for the adjustment of profile likelihoods. Journal of the Royal Statistical Society. Series B 52, 325–344.
Molenberghs, G., Verbeke, G., 2004. Meaningful statistical model formulations for repeated measures. Statistica Sinica 14, 989–1020.
Mukherjee, B., Ahn, J., Liu, I., Rathouz, P.J., Sanchez, B.N., 2008. Fitting stratified proportional odds models by amalgamating conditional likelihoods. Statistics in Medicine 27, 4950–4971.
Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam.
Neyman, J., Scott, E.L., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.
Rasch, G., 1961, On general laws and the meaning of measurement in psychology. In: Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 321–333.
White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–26.
Wooldridge, J.M., 2000. A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. Economics Letters 68, 245–250.

Journal of Statistical Software

June 2017, Volume 78, Issue 7. doi: 10.18637/jss.v078.i07

cquad: An R and Stata Package for Conditional Maximum Likelihood Estimation of Dynamic Binary Panel Data Models

Francesco Bartolucci
University of Perugia

Claudia Pigini
Marche Polytechnic University

Abstract

We illustrate the R package **cquad** for conditional maximum likelihood estimation of the quadratic exponential (QE) model proposed by Bartolucci and Nigro (2010) for the analysis of binary panel data. The package also allows us to estimate certain modified versions of the QE model, which are based on alternative parametrizations, and it includes a function for the pseudo-conditional likelihood estimation of the dynamic logit model, as proposed by Bartolucci and Nigro (2012). We also illustrate a reduced version of this package that is available in Stata. The use of the main functions of this package is based on examples using labor market data.

Keywords: dynamic logit model, pseudo maximum likelihood estimation, quadratic exponential model, state dependence.

1. Introduction

With the growing number of panel datasets available to practitioners and the recent development of related statistical and econometric models, ready-to-use software to estimate non-linear models for binary panel data is now essential in applied research. In particular, the panel structure allows for formulations that include both unobserved heterogeneity (i.e., time-constant individual intercepts) and the lagged response variable, which accounts for the so-called state dependence (i.e., how the experience of a certain event affects the probability of experiencing the same event in the future), as defined in Heckman (1981a).

A simple and, at the same time, interesting approach for the analysis of binary panel data is based on the dynamic logit (DL) model, which includes individual-specific intercepts and state dependence. The estimation of such a model may be based either on a random-effects or on a fixed-effects formulation. In the first case, individual intercepts are treated as random parameters while, in the second, each intercept is considered as a fixed parameter to be estimated. The fixed-effects approach attracts considerable attention as it requires a reduced amount of assumptions with respect to the random-effects formulation, based on the independence between the individual unobserved effects and the observable covariates, and on the normality assumption.

For the static fixed-effects logit model (i.e., the DL model without the lagged response variable among the covariates), it is possible to eliminate the individual intercepts by conditioning on simple sufficient statistics (Andersen 1970; Chamberlain 1980). In general, the estimator based on this method is known as conditional maximum likelihood (CML) estimator. The full DL model, however, does not admit simple sufficient statistics for the individual intercepts and, therefore, cannot be estimated by CML in a simple way as the static logit model.

The drawback described above is overcome by Bartolucci and Nigro (2010), who develop a model for the analysis of dynamic binary panel data models based on a Quadratic Exponential (QE) formulation (Cox 1972), which has the advantage of admitting sufficient statistics for the unobserved heterogeneity parameters. Therefore, the model parameters can easily be estimated by the CML method. Recently, further extensions to the approach of Bartolucci and Nigro (2010) have also been proposed. In particular, Bartolucci and Nigro (2012) propose a QE model that closely approximates the DL model. Finally, Bartolucci, Nigro, and Pigini (2017) derive a test for state dependence that is more powerful than the one based on the standard QE model.

In this paper we illustrate **cquad** (Bartolucci and Pigini 2017), which is a comprehensive R (R Core Team 2017) package for the CML estimation of fixed-effects binary panel data models. In particular, **cquad** contains functions for the estimation of the static logit model (Chamberlain 1980), and of the dynamic QE models recently proposed by Bartolucci and Nigro (2010, 2012) and Bartolucci *et al.* (2017). A version of the R package **cquad**, including its main functionalities, is also available for Stata (StataCorp. 2015; Bartolucci 2015) and is illustrated here.

As it implements fixed-effects estimators of non-linear panel data models for binary dependent variables, **cquad** complements the existing array of R packages for panel data econometrics. Above all, it is closely related to the **plm** package (see Croissant and Millo 2008), which provides a wide set of functions for the estimation of linear panel data models for both static and dynamic formulations. In addition, **cquad** shares with **plm** the peculiarities of the data frame structure, of the formula supplied to `model.matrix`, and of the object class `panelmodel`. **cquad** is also related to package **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and R Core Team 2017), which implements non-linear mixed-effects models that can be estimated with longitudinal data.

The Stata module **cquad** represents an addition to the many existing commands and modules for panel data econometrics available in this software, such as `xtreg` and `xtabond2` for linear models, and it complements the available routine for the CML and ML estimation of the static logit model, namely the native `xtlogit`. In addition, it relates to the routines and modules for the estimation of static random-effects binary panel data models, such as the built-in `xtprobit` and the module **gllamm** (2011) for the estimation for generalized linear mixed models (see Rabe-Hesketh, Skrondal, and Pickles 2005), and the implementation of dynamic models, in the modules **redprob** and **redpace** (see Stewart 2006).

Finally, a package for the estimation of binary panel data models with similar functionalities is the **DPB** function package for **gretl** (see Lucchetti and Pigini 2015, for details), which implements the CML estimator for the QE model by Bartolucci and Nigro (2010). A related package, which however uses a different approach for parameter estimation, is the R package **panelMPL** described in Bartolucci, Bellio, Salvan, and Sartori (2016).

The paper is organized as follows. In the next section we briefly review the basic definition of the DL model and of the different versions of the QE model here considered. We also briefly review CML and pseudo-CML estimation of the models. Then, in Section 3 we describe the main functionalities of package **cquad** for R and the corresponding module for Stata. Finally, the illustration of the packages by examples is provided in Section 4.

For the purpose of describing **cquad** functionalities, we use data on unionized workers extracted from the U.S. National Longitudinal Survey of Youth. In particular, to illustrate the R package, we use the same data as in Wooldridge (2005), whereas for the Stata module we employ similar data already available in the Stata repository.

2. Preliminaries

We consider a binary panel dataset referred to a sample of n units observed at T consecutive time occasions. We adopt a common notation in which y_{it} is the response variable for unit i at occasion t, with $i = 1, \ldots, n$ and $t = 1, \ldots, T$, and $\boldsymbol{x}_{it}$ is the corresponding column of covariates. In the following we first describe the CML method applied to the logit model, then we illustrate the DL and QE models for the analysis of dynamic binary panel data models and inference based on the CML method.

2.1. Conditional maximum likelihood estimation

In order to provide an outline of the CML method by Andersen (1970), in the following we describe the derivation of the conditional likelihood for the static logit model (Chamberlain 1980), which will be the basic framework for the QE models described later in this section.

Consider the static logit formulation based on the assumption

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

where α_i is the individual specific intercept and vector $\boldsymbol{\beta}$ collects the regression parameters associated with the explanatory variables $\boldsymbol{x}_{it}$. For the joint probability of $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})^\top$, this model implies that

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i) = \frac{\exp\left(\alpha_i y_{i+}\right) \exp\left(\sum_t y_{it} \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)}{\prod_t \left[1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)\right]},$$

where the sum $\sum_t$ and product $\prod_t$ range over $t = 1, \ldots, T$ and $y_{i+} = \sum_t y_{it}$ is called the *total score*.

It can be shown that y_{i+} is a sufficient statistic for the individual intercepts α_i (Andersen 1970). Consequently, the joint probability of $\boldsymbol{y}_i$, conditional on y_{i+}, does not depend on α_i. In fact, we have

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

where the denominator is the sum of the probabilities of observing each possible vector configuration of binary responses $\boldsymbol{z} = (z_1, \dots, z_T)^\top$ such that $z_+ = y_{i+}$, where $z_+ = \sum_t z_t$, that is,

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i+}) = \frac{p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i)}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} p(\boldsymbol{z}|\alpha_i, \boldsymbol{X}_i)},$$

with

$$p(\boldsymbol{z}|\alpha_i, \boldsymbol{X}_i) = \frac{\exp(\alpha_i z_+)\exp\left(\sum_t z_t \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)}{\prod_t \left[1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)\right]}.$$

Therefore, the conditional distribution of the vector of responses $\boldsymbol{y}_i$ is

$$\begin{aligned} p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i+}) &= \frac{\exp(\alpha_i y_{i+})\exp\left(\sum_t y_{it}\boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)}{\prod_t \left[1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)\right]} \frac{\prod_t \left[1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)\right]}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp(\alpha_i z_+)\exp\left(\sum_t z_+ \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)} \\ &= \frac{\exp\left(\sum_t y_{it}\boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp\left(\sum_t z_t \boldsymbol{x}_{it}^\top \boldsymbol{\beta}\right)} = p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}), \end{aligned}$$

where the individual intercepts α_i have been canceled out.

The conditional log-likelihood based on the above distribution can be written as

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

where the indicator function $\mathbf{I}(\cdot)$ is introduced to take into account that observations whose total score is 0 or T do not contribute to the likelihood. This conditional log-likelihood can be maximized with respect to $\boldsymbol{\beta}$ by a Newton-Raphson algorithm, obtaining the CML estimator $\hat{\boldsymbol{\beta}}$. Expressions for the score vector and information matrices can be derived using the standard theory on the regular exponential family (Barndorff-Nielsen 1978).

2.2. Dynamic logit model

The DL model (Hsiao 2005) represents an interesting dynamic approach for binary panel data as it includes, apart from the observable covariates, both individual specific intercepts and the lagged response variable. Its formulation is a simple extension of Equation 1 with also $y_{i,t-1}$ in the set of covariates.

For a sequence of binary responses y_{it}, $t = 1, \dots, T$, referred to the same unit i, and the corresponding covariate vectors $\boldsymbol{x}_{it}$, the conditional distribution of a single response is

$$p(y_{it}|\alpha_i, \boldsymbol{X}_i, y_{i0}, \dots, y_{i,t-1}) = \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + y_{i,t-1}\gamma)]}{1 + \exp(\alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + y_{i,t-1}\gamma)}, \tag{2}$$

where γ is the regression coefficient for the lagged response variable measuring the true state dependence.

The inclusion of the individual intercept α_i for the unobserved heterogeneity in a dynamic model raises the so-called "initial conditions" problem (Heckman 1981b), which concerns the correlation between time-invariant effects and the initial realization of the outcome, y_{i0}.

However, with a fixed-effects approach, individual unobserved effects are treated as fixed parameters and the initial observation can be considered as given. The distribution of the vector of responses $\boldsymbol{y}_i$ conditional on y_{i0} is

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp\left(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\beta} + y_{i*}\gamma\right)}{\prod_t\left[1+\exp\left(\alpha_i + \boldsymbol{x}_{it}^\top\boldsymbol{\beta} + y_{i,t-1}\gamma\right)\right]}, \tag{3}$$

where $y_{i*} = \sum_t y_{i,t-1}y_{it}$.

Differently from the static logit model in Equation 1, the full DL model does not admit sufficient statistics for the individual parameters α_i. Therefore, CML inference is not viable in a simple form, but can only be derived in the special case of $T = 3$ and in absence of explanatory variables (Chamberlain 1985). Honoré and Kyriazidou (2000) extend this approach to include covariates in the regression model, so that parameters are estimated by CML on the basis of a weighted conditional log-likelihood. However, their approach presents some limitations; mainly, discrete covariates cannot be included in the model specification and, although the estimator is consistent, its rate of convergence to the true parameter value is slower than $\sqrt{n}$.

2.3. Quadratic exponential models

The shortcomings of the fixed-effects DL model can be overcome by the approximating QE model defined in Bartolucci and Nigro (2010), based on the family of distributions for multivariate binary data formulated by Cox (1972). The QEext model directly formulates the conditional distribution of $\boldsymbol{y}_i$ as follows:

$$p(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp\left[y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\eta}_1 + y_{iT}\left(\phi + \boldsymbol{x}_{iT}^\top\boldsymbol{\eta}_2\right) + y_{i*}\psi\right]}{\sum_{\boldsymbol{z}}\exp[z_+\delta_i + \sum_t z_t\boldsymbol{x}_{it}^\top\boldsymbol{\eta}_1 + z_T\left(\phi + \boldsymbol{x}_{iT}^\top\boldsymbol{\eta}_2\right) + z_{i*}\psi]}, \tag{4}$$

where δ_i is the individual specific intercept, $\sum_{\boldsymbol{z}}$ ranges over the possible binary response vectors $\boldsymbol{z}$, and $z_{i*} = y_{i0}z_1 + \sum_{t>1} z_{t-1}z_t$. The parameter ψ measures the true state dependence and vector $\boldsymbol{\eta}_1$ collects the regression parameters associated with the covariates. Here we consider ϕ and $\boldsymbol{\eta}_2$ as nuisance parameters. We refer the reader to Bartolucci and Nigro (2010) for the discussion on the interpretation of these parameters.

The QE model allows for state dependence and unobserved heterogeneity, other than the effect of observable covariates, some of which may be discrete. Moreover, it shares several properties with the DL model:

1. for $t = 2, \ldots, T$, y_{it} is conditionally independent of $y_{i0}, \ldots, y_{i,t-2}$, given $\boldsymbol{X}_i, y_{i,t-1}$, and α_i or δ_i, under both models;

2. for $t = 1, \ldots, T$, the conditional log-odds ratio for $(y_{i,t-1}, y_{it})$ is constant:

$$\log\frac{p(y_{it}=1|\delta_i, \boldsymbol{X}_i, y_{i,t-1}=1)p(y_{it}=0|\delta_i, \boldsymbol{X}_i, y_{i,t-1}=0)}{p(y_{it}=0|\delta_i, \boldsymbol{X}_i, y_{i,t-1}=1)p(y_{it}=1|\delta_i, \boldsymbol{X}_i, y_{i,t-1}=0)} = \psi,$$

while in the DL model it is constant and equal to γ.

Differently from the DL model, the QE model does admit a sufficient statistic for the individual intercepts δ_i. The parameters for the unobserved heterogeneity are removed by condition on the total score y_{i+}. In particular, following the same derivations as in Section 2.1, we obtain:

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp[\sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\eta}_1 + y_{iT}(\phi + \boldsymbol{x}_{iT}^\top\boldsymbol{\eta}_2) + y_{i*}\psi]}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp[\sum_t z_t\boldsymbol{x}_{it}^\top\boldsymbol{\eta}_1 + z_T(\phi + \boldsymbol{x}_{iT}^\top\boldsymbol{\eta}_2) + z_{i*}\psi]}. \tag{5}$$

The parameter vector $\boldsymbol{\theta} = (\boldsymbol{\eta}_1^\top, \phi, \boldsymbol{\eta}_2^\top, \psi)^\top$ can be estimated by maximizing the conditional log-likelihood based on Equation 5, that is,

$$\ell(\boldsymbol{\theta}) = \sum_i \mathbf{I}(0 < y_{i+} < T)\log p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}).$$

As for the static logit model, this maximization may simply be performed by a Newton-Raphson algorithm, and the resulting estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\eta}}_1^\top, \hat{\phi}, \hat{\boldsymbol{\eta}}_2^\top, \hat{\psi})^\top$ is $\sqrt{n}$-consistent and has asymptotic normal distribution. For the derivation of the score vector and the information matrix and of the expression for the standard errors, we refer the reader to Bartolucci and Nigro (2010).

A simplified version of the QEext model can be derived by assuming that the regression parameters are equal for all time occasions. The joint probability of the individual outcomes of this model, which we will refer to as QEbasic hereafter, is expressed as

$$p_b(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\eta} + y_{i*}\psi)}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t\boldsymbol{x}_{it}^\top\boldsymbol{\eta} + z_{i*}\psi)}. \tag{6}$$

In the same way as for the QEext model, a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta} = (\boldsymbol{\eta}^\top, \psi)^\top$ can be obtained by maximizing the conditional log-likelihood based on (6) by a Newton-Raphson algorithm.

Finally, Bartolucci *et al.* (2017) introduce a test for state dependence based on a modified version of the QEbasic model, named QEequ hereafter. The joint probability of $\boldsymbol{y}_i$ is defined as

$$p_e(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\eta} + \tilde{y}_{i*}\psi)}{\sum_{\boldsymbol{z}} \exp(z_+\delta_i + \sum_t z_t\boldsymbol{x}_{it}^\top\boldsymbol{\eta} + \tilde{z}_{i*}\psi)}, \tag{7}$$

where $\tilde{y}_{i*} = \sum_t \mathbf{I}\{y_{it} = y_{i,t-1}\}$ and $\tilde{z}_{i*} = \mathbf{I}\{z_1 = y_{i0}\} + \sum_{t>1}\mathbf{I}\{z_t = z_{t-1}\}$. The difference with the QE models described earlier is in how the association between the response variables is formulated: this modified version is based on the statistic $\tilde{y}_{i*}$ that, differently from y_{i*}, is equal to the number of consecutive pairs of outcomes that are equal each other, regardless of whether they are 0 or 1. This allows us to use a larger set of information with respect to the QEext and QEbasic in testing for state dependence.

Conditioning on the total score y_{i+}, the expression for the joint probability becomes

$$p_e(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\eta} + \tilde{y}_{i*}\psi)}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t\boldsymbol{x}_{it}^\top\boldsymbol{\eta} + \tilde{z}_{i*}\psi)}. \tag{8}$$

In the same way as for the QEext and QEbasic model, $\boldsymbol{\theta} = (\boldsymbol{\eta}^\top, \psi)^\top$ can be consistently estimated by CML and, in particular, by maximizing the conditional log-likelihood based on (8), obtaining $\hat{\boldsymbol{\theta}}_e = (\hat{\boldsymbol{\eta}}_e, \hat{\psi}_e)$.

Once the parameters in Equation 7 are estimated, a t-statistic for $H_0 : \psi = 0$ is

$$W = \frac{\hat{\psi}_e}{\mathrm{se}(\hat{\psi}_e)}, \tag{9}$$

where $\mathrm{se}(\cdot)$ is the standard error derived using the sandwich estimator; see Bartolucci *et al.* (2017) for the complete derivation of score, information matrix, and variance-covariance matrix.

Under the DL model, and provided that the null hypothesis $H_0 : \gamma = 0$ holds, the test statistic W has asymptotic standard normal distribution as $n \to \infty$. If $\gamma \neq 0$, W diverges to $+\infty$ or $-\infty$ according to whether γ is positive or negative.

2.4. Pseudo-conditional maximum likelihood estimation

In order to estimate the structural parameters of the DL model, Bartolucci and Nigro (2012) propose a pseudo-CML estimator based on approximating this model by a QE model of the type described in Section 2.3. The proposed approximating model also has the advantage of admitting a simple sufficient statistic for each individual intercept and its parameters share the same interpretation as the true DL model.

The approximating model is derived from a linearization of the log-probability of the DL model defined in Equation 3, that is,

$$\log p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\beta} + y_{i*}\gamma - \sum_t \log[1 + \exp(\alpha_i + \boldsymbol{x}_{it}^\top\boldsymbol{\beta} + y_{i,t-1}\gamma)].$$

The non-linear component is approximated by a first-order Taylor series expansion around $\alpha_i = \bar{\alpha}$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, and $\gamma = 0$:

$$\begin{aligned}\sum_t \log[1 + \exp(\alpha_i + \boldsymbol{x}_{it}^\top\boldsymbol{\beta} + y_{i,t-1}\gamma)] \quad &\approx \quad \sum_t \Big\{\log\Big[1 + \exp\Big(\bar{\alpha}_i + \boldsymbol{x}_{it}^\top\bar{\boldsymbol{\beta}}\Big)\Big] \\ + \bar{q}_{it}\Big[\alpha_i - \bar{\alpha}_i + \boldsymbol{x}_{it}^\top(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\Big]\Big\} \quad &+ \quad \bar{q}_{i1}y_{i0}\gamma + \sum_{t>1}\bar{q}_{it}y_{i,t-1}\gamma,\end{aligned}$$

where $\bar{q}_{it} = \exp(\bar{\alpha}_i + \boldsymbol{x}_{it}^\top\bar{\boldsymbol{\beta}})/[1 + \exp(\bar{\alpha}_i + \boldsymbol{x}_{it}^\top\bar{\boldsymbol{\beta}})]$. Under this approximating model, referred to QEpseudo hereafter, the joint probability of $\boldsymbol{y}_i$ is

$$p_p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\beta} - \sum_t \bar{q}_{it}y_{i,t-1}\gamma + y_{i*}\gamma)}{\sum_{\boldsymbol{z}} \exp(z_+\alpha_i + \sum_t z_t\boldsymbol{x}_{it}^\top\boldsymbol{\beta} - \sum_t \bar{q}_{it}z_{i,t-1}\gamma + z_{i*}\gamma)}. \tag{10}$$

Given α_i and $\boldsymbol{X}_i$, the above model corresponds to a quadratic exponential model (Cox 1972) with second-order interactions equal to γ, when referred to consecutive response variables, and to 0 otherwise.

Under the approximating model, each y_{i+} is a sufficient statistic for the incidental parameter α_i. By conditioning on the total scores, the joint probability of $\boldsymbol{y}_i$ becomes:

$$p_p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_t y_{it}\boldsymbol{x}_{it}^\top\boldsymbol{\beta} - \sum_t \bar{q}_{it}y_{i,t-1}\gamma + y_{i*}\gamma)}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t\boldsymbol{x}_{it}^\top\boldsymbol{\beta} - \sum_t \bar{q}_{it}z_{i,t-1}\gamma + z_{i*}\gamma)}, \tag{11}$$

where the individual intercepts α_i cancel out.

A pseudo-CML estimator based on the approximating model described in Equation 11 is introduced by Bartolucci and Nigro (2012). The estimator is based on the following two-step procedure:

1. A preliminary estimate of the regression parameter $\boldsymbol{\beta}$, $\tilde{\boldsymbol{\beta}}$, is computed by maximizing the conditional log-likelihood of the static logit model described in Section 2.1. In addition, the probabilities $\bar{q}_{it}$, for $i = 1, \ldots, n$ and $t = 2, \ldots, T$, are computed with $\bar{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ and $\bar{\alpha}_i$ equal to its maximum likelihood estimate under the static logit model.

2. The parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \gamma)^\top$ is estimated by maximizing the conditional log-likelihood

$$\ell_p(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \sum_i \mathbf{I}\{0 < y_{i+} < T\} \log p_p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}).$$

The maximization of $\ell_p(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}})$ is possible by a simple Newton-Raphson algorithm, resulting in the pseudo-CML estimator $\hat{\boldsymbol{\theta}}_p = (\hat{\boldsymbol{\beta}}_p^\top, \hat{\gamma}_p)^\top$ of the structural parameters of the DL model. For asymptotic results and computation of standard errors we refer the reader to Bartolucci and Nigro (2012).

3. Package description

Here we describe the main functionalities of the R package **cquad** and then the corresponding commands of the **cquad** module implemented in Stata.

3.1. The R package

The cquad *interface*

Package **cquad** includes several functions, the majority of which are called by the main interface `cquad`. The first argument of the `cquad` function is a formula that shares the same syntax with that of the **plm** package. For instance, using the sample data on unionized workers, `Union.RData`, a simple function call is

```
R> cquad(union ~ married, Union)
```

where the dependent variable must be a numeric binary vector. In general, as in `plm` and differently from `lm`, the formula can also recognize the operators `lag`, `log`, and `diff` that can be supplied directly without additional transformations of the covariates.

The second argument supplied to `cquad` is the data frame. As in **plm**, the data must have a panel structure, that is the data frame has to contain an individual identifier and a time variable as the first two columns. For instance, the data frame `Union` has the following structure:

```
R> head(Union[c(1, 2)])
```

```
  nr year
1 13 1980
2 13 1981
3 13 1982
4 13 1983
5 13 1984
6 13 1985
```

where `nr` is the individual identifier and `year` provides the time variable. As `Union` already has a panel structure, `cquad` can be called directly. Differently, if the dataset does not contain the individual and time indicators, `cquad` sets the panel structure and creates automatically the first two variables, provided `index` is supplied, namely the number of cross-section observations in the data. As an example, the dataset `Wages`, supplied by **plm** and containing 595 individuals observed over 7 periods, does not have a panel structure, which however is created by `cquad` as follows:

```
R> cquad(union2 ~ married, Wages, index = 595)
```

Package **cquad** uses the same function as **plm** to impose the panel structure on a data frame, called `plm.data`. Indeed, this function can also be used to set the panel structure to the data frame, which can then be supplied to `cquad` without the index argument. For instance:

```
R> Wages <- plm.data(Wages, 595)
```

produces

```
R> head(Wages)
```

```
  id time exp wks bluecol ind south smsa married  sex union ed black   lwage
1  1    1   3  32      no   0   yes   no     yes male    no  9    no 5.56068
2  1    2   4  43      no   0   yes   no     yes male    no  9    no 5.72031
3  1    3   5  40      no   0   yes   no     yes male    no  9    no 5.99645
4  1    4   6  39      no   0   yes   no     yes male    no  9    no 5.99645
5  1    5   7  42      no   1   yes   no     yes male    no  9    no 6.06146
6  1    6   8  35      no   1   yes   no     yes male    no  9    no 6.17379
```

where the factors `id` and `time` have been created and added to the data frame.

In the examples above, both data frames refer to balanced panels. Nevertheless, **cquad** also handles unbalanced panels.

Each of the models described in Section 2 is estimated by **cquad** by supplying a dedicated string to the function argument `model`. In particular, we can estimate:

- the fixed-effects static logit model by Chamberlain (1980) (`model = "basic"`, default);
- the simplified QE model, QEbasic (`model = "basic", dyn = TRUE`);
- the QEext model proposed by Bartolucci and Nigro (2010) (`model = "extended"`);

- the modified version of the QE model, QEequ proposed in Bartolucci *et al.* (2017) (`model = "equal"`);
- the pseudo-CML estimation of the DL model based on the approach of Bartolucci and Nigro (2012) (`model = "pseudo"`).

As an optional argument, the `cquad` function can also be supplied with an n-dimensional vector of individual weights; the default value is `rep(1, n)`.

The results of the calls to `cquad` are stored in an object of class `panelmodel`. The returned object shares only some elements with a `panelmodel` object and contains additional ones due to the peculiarities of CML inference.

The elements in common with the object `panelmodel`, as described in **plm**, are `coefficients`, `vcov`, and `call`. The vector `coefficients` contains the estimates of: the k-dimensional vector $\boldsymbol{\beta}$, for the static logit; the $(k+1)$-dimensional vector $\boldsymbol{\theta} = (\boldsymbol{\eta}^\top, \psi)^\top$ for the dynamic models QEbasic, the conditional probability of which is defined in Equation 6, and QEequ in Equation 7, respectively; the $(2k+2)$-dimensional vector $\boldsymbol{\theta} = (\boldsymbol{\eta}_1^\top, \phi, \boldsymbol{\eta}_2^\top, \psi)^\top$ for the QEext model in Equation 4; the $(k+1)$-dimensional vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \gamma)^\top$ in Equation 10 for the pseudo-CML estimator of the DL model. The matrix `vcov` contains the corresponding asymptotic variance-covariance matrix for the parameter estimates. Finally, `call` contains the function call to the sub-routines required to fit each model, namely `cquad_basic`, `cquad_ext`, `cquad_equ`, or `cquad_pseudo`.

The output of `cquad` does not provide fitted values nor residuals: as discussed in Section 2, the CML estimation approach is based on eliminating the individual intercepts in each model, and this does not allow for the computation of predicted probabilities. Similarly, residuals are not a viable tool for standard inference. On the other hand, we supply the object with estimated quantities useful for inference and diagnostics within the CML estimation approach.

The asymptotic standard errors associated with the estimated coefficients are collected in the vector `se` and the robust standard errors (White 1980) in vector `ser`. For the pseudo-CML estimator, the standard errors contained in the vector `ser` are corrected for the presence of estimated regressors (see Bartolucci and Nigro 2012, for the detailed derivation of the two-step variance-covariance matrix). The function output also provides the matrix `scv` containing the individual scores and the matrix `J` containing the Hessian of the log-likelihood function. In addition, `cquad` returns the conditional log-likelihood at convergence (`lk`) for each of the fitted models. Finally, it contains the n-dimensional vector `Tv` of the number of observations for each unit.

Simulate data from the DL model

Package **cquad** also contains function `sim_panel_logit`, which allows the user to generate a binary vector from a DL data generating process. This function requires in input the list of unit identifiers in the panel, which are collected in vector `id` having length equal to the overall number of observations $n \times T = r$. As other inputs, the function requires the n-dimensional vector of the individual specific intercepts that must be somehow generated, for instance drawing them from a standard normal distribution, and the matrix of covariates (if they exist) that has dimension $r \times k$, where k is the number of covariates. Each row of this matrix contains a vector of covariates $\boldsymbol{x}_{it}$ arranged according to vector `id`. Finally, in input the function requires the vector of structural parameters, denoted by `eta`, that is, $\boldsymbol{\beta}$ for the static logit model and $(\boldsymbol{\beta}^\top, \gamma)^\top$ for the DL model; the model of interest is specified by the optional argument `dyn`.

As output values, function `sim_panel_logit` returns a list containing two vectors, `pv` and `yv`. The first contains the success probability computed according to the DL model corresponding to each row of matrix `X` and accounting for the corresponding individual intercept in `al`. Vector `yv` contains the binary variable which is randomly drawn from this distribution.

3.2. The Stata module

The **cquad** module in Stata consists of four Mata routines for the estimation by CML of the QE models described in Section 2.3. It contains four commands with the syntax

`cquad`*cmd depvar* `id` [*indepvars*]

where *cmd* has to be substituted with the string corresponding to the type of model to be estimated. In particular:

- `cquadext` fits the QEext model of Bartolucci and Nigro (2010) defined in Equation 4;
- `cquadbasic` estimates the parameters of the simplified QE model, QEbasic, the conditional probability of which is defined in Equation 6. Differently from the R package, `cquadbasic` fits only the dynamic QE model, as the static logit model can estimated by `xtlogit`;
- `cquadequ` fits the modified QE model defined in Equation 7 proposed by Bartolucci *et al.* (2017);
- `cquadpseudo` fits the pseudo-CML estimator proposed by Bartolucci and Nigro (2012) for the parameters in Equation 10.

In addition, *depvar* is the series containing the binary dependent variable, and `id` is the variable containing the list of reference units uniquely identifying individuals in the panel dataset. Optionally a list of covariates [*indepvars*] can be supplied.

The four commands return an `eclass` object with the estimation results. Scalar `e(lk)` contains the final conditional log-likelihood and macro `e(cmd)` holds the function call. Moreover, matrix `e(be)` contains the estimated coefficients and it is of dimension $(2k + 2) \times 1$ for `cquadext`, or of dimension $(k + 1) \times 1$ for `cquadbasic`, `cquadequ`, and `cquadpseudo`. Matrices `e(se)` and `e(ser)` contain the corresponding estimated asymptotic and robust standard errors, respectively. Finally, matrices `e(tstat)` and `e(pv)` collect the t test statistics and the corresponding p values.

4. Examples

In the following we illustrate package **cquad** by means of three applications. In particular, we show how to compute the CML estimators for the QE models and the pseudo-CML estimator in R and Stata using longitudinal data on unionized workers extracted from the U.S. National Longitudinal Survey of Youth, which has been employed in several applied works to illustrate dynamic binary panel data models (Wooldridge 2005; Stewart 2006; Lucchetti and Pigini 2015). Moreover, we propose a simulation example using `sim_panel_logit` provided in the R package.

4.1. Use of the `Union` dataset in R

To illustrate the R package, we use the dataset employed in Wooldridge (2005) and available in the *Journal of Applied Econometrics* data archive. The dataset is referred to 545 male workers interviewed for eight years, from 1980 to 1987. Similarly to the empirical application in Wooldridge (2005), the variables relevant to our example are a binary variable equal to 1 if the worker's wage is set by a union, which will be used as the dependent variable, and a binary variable describing his marital status, used as covariate. The original dataset also contains information on the race and years of schooling, which however cannot be employed in our example since they are time-invariant:

```
  nr year black married educ union
1 13 1980     0       0   14     0
2 13 1981     0       0   14     1
3 13 1982     0       0   14     0
4 13 1983     0       0   14     0
5 13 1984     0       0   14     0
6 13 1985     0       0   14     0
```

Notice that the panel structure required by **cquad** is already imposed.

Then, in order to fit the static logit model to this data by the CML method, we call `cquad` with the following syntax

```
R> out1 <- cquad(union ~ married + year, Union)
```

This estimates a logit model with `union` as the dependent variable and `married` and time dummies as covariates, obtaining the following output

```
Balanced panel data
 |--------------|--------------|--------------|
 |   iteration  |      lk      |    lk-lko    |
 |--------------|--------------|--------------|
 |            1 |     -740.781 |          Inf |
 |            2 |      -732.45 |       8.3312 |
 |            3 |     -732.445 |   0.00539603 |
 |            4 |     -732.445 |  9.75388e-09 |
 |--------------|--------------|--------------|
```

Then, using command `summary(out1)`, we obtain:

```
Call:
cquad_basic(id = id, yv = yv, X = X, w = w, dyn = dyn)

Log-likelihood:
```

```
-732.4449
                est.      s.e.       t-stat     p-value
married   0.298326773 0.1708112  1.746529038 0.080719066
year1981 -0.061754846 0.2061185 -0.299608423 0.764475859
year1982  0.000927442 0.2069901  0.004480611 0.996425002
year1983 -0.155186804 0.2117482 -0.732883615 0.463629417
year1984 -0.107846793 0.2137133 -0.504633157 0.613816517
year1985 -0.442338283 0.2189339 -2.020419690 0.043339873
year1986 -0.608785100 0.2222082 -2.739705640 0.006149423
year1987 -0.015457650 0.2180398 -0.070893720 0.943482341
```

The output of `summary` displays the function call, the value of the log-likelihood at convergence, and the estimated coefficients with the corresponding asymptotic standard errors and t test results. Notice that including variable `year` among the covariates in the formula leads `cquad` to the automatic inclusion of the time dummies in the model specification, except for `year1980` due to collinearity, even though variable `year` is numeric in the original data frame:

```
R> str(Union$year)
```

```
int [1:4360] 1980 1981 1982 1983 1984 1985 1986 1987 1980 1981 ...
```

This happens because `cquad` recognizes the second variable in the data frame as the time variable, and with the call to `plm.data` and `model.matrix` the numeric time variable is transformed into a factor.

To estimate the dynamic specification of the QEbasic model, `cquad` needs to be called with the `dyn = TRUE` option. In addition, as we are working with a balanced panel, an additional time dummy must be excluded because the lag of the dependent variable is included in the conditioning set and the initial time occasion is lost. In this case, we perform this operation outside the `cquad` interface

```
R> year2 <- Union$year
R> year2[year2 == 1980 | year2 == 1981] <- 0
R> year2 <- as.factor(year2)
R> out2 <- cquad(union ~ married + year2, Union, dyn = TRUE)
R> summary(out2)
```

In the code above, we store the numeric time variable from the original data frame in `year2`; then, we set the variable to 0 for two of its values, as we loose one time occasion due to the dynamic specification and one time effect due to the collinearity of the remaining dummies. In order to estimate the model with time dummies, we need to convert `year2` into a factor: `cquad` will not recognize `year2` as the time variable since it is not in the data frame. If instead we leave `year` in the formula, a warning message is given after convergence and the results are obtained using the generalized inverse of the Hessian matrix.

The estimation output produced by the above command lines is (iteration logs are omitted from the output below)

```
Call:
cquad_basic(id = id, yv = yv, X = X, w = w, dyn = dyn)
```

```
Log-likelihood:
-505.514
                  est.      s.e.     t-stat      p-value
married    0.13404719 0.1868762  0.7173047 0.4731861145
year21982  0.09160286 0.2441350  0.3752140 0.7075013011
year21983 -0.09896744 0.2258889 -0.4381245 0.6612960556
year21984  0.09917729 0.2254660  0.4398770 0.6600262259
year21985 -0.27210110 0.2309277 -1.1782956 0.2386787776
year21986 -0.52465221 0.2328383 -2.2532900 0.0242408710
year21987  0.81055556 0.2265106  3.5784449 0.0003456447
y_lag      1.47082575 0.1528797  9.6208037 0.0000000000
```

Although `cquad` with `model = "basic"` (default) and `dyn = TRUE` fits the simplified version of the QE model (i.e., QEbasic), which approximates the true DL model, the obtained results are in line with the findings on the probability of participating in a union under dynamic models: there is a positive and significant correlation with the lagged dependent variable ($\psi = 1.471$), and the effect of `married` is not statistically significant.

To fit the QEext model, we need to further exclude the last time value (i.e., 1987): since there is an intercept term ϕ in Equation 5, the effect associated with the last time dummy is not identified with balanced panels:

```
R> year3 <- Union$year
R> year3[year3 == 1980 | year3 == 1981 | year3 == 1987] <- 0
R> year3 <- as.factor(year3)
R> out3 <- cquad(union ~ married + year3, Union, model = "extended")
```

By typing `summary(out3)` we obtain

```
Call:
cquad_ext(id = id, yv = yv, X = X, w = w)

Log-likelihood:
-504.2864
                     est.      s.e.      t-stat    p-value
married        0.01958449 0.2008834  0.09749182 0.92233583
year31982      0.09808421 0.2442447  0.40158167 0.68799192
year31983     -0.08051308 0.2262232 -0.35590102 0.72191469
year31984      0.12301583 0.2259423  0.54445680 0.58612717
year31985     -0.24494702 0.2314885 -1.05813907 0.28999205
year31986     -0.48914076 0.2339525 -2.09076982 0.03654870
int            0.51995850 0.2952783  1.76091005 0.07825363
diff.married   0.51942916 0.3328688  1.56046215 0.11865071
y_lag          1.47056206 0.1530829  9.60631199 0.00000000
```

where the additional `int` and `diff.` variables represent ϕ and $\boldsymbol{\eta}_2$ in Equation 4, respectively.

Similarly, to fit the QEequ model defined in Equation 7 and display the results, the command lines are as follows:

```
R> out4 <- cquad(union ~ married + year2, Union, model = "equal")
R> summary(out4)
```

which returns

```
Call:
cquad_equ(id = id, yv = yv, X = X, w = w)

Log-likelihood:
-505.514
                 est.       s.e.     t-stat    p-value
married    0.13404719 0.18687622  0.7173047 0.47318611
year21982  0.09160286 0.24413496  0.3752140 0.70750130
year21983 -0.09896744 0.22588886 -0.4381245 0.66129606
year21984  0.09917729 0.22546598  0.4398770 0.66002623
year21985 -0.27210110 0.23092771 -1.1782956 0.23867878
year21986 -0.52465221 0.23283830 -2.2532900 0.02424087
year21987  0.07514269 0.21352948  0.3519078 0.72490741
y_lag      0.73541287 0.07643986  9.6208037 0.00000000
```

Notice that there is a marked difference in the estimated coefficient associated with the lagged dependent variable. In model QEequ, the association between y_{it} and $y_{i,t-1}$ is different from that of the standard formulation of the QE model so as to exploit more information in testing for state dependence (see Section 2.3). Indeed, the t test statistic associated with `y_lag` is referred to the test for state dependence described in Equation 9.

In order to fit the pseudo-CML model, `cquad` needs to be called with `model = "pseudo"`:

```
R> out5 <- cquad(union ~ married + year2, Union, model = "pseudo")
```

that produces the output

```
First step estimation
Balanced panel data
 |--------------|--------------|--------------|
 |   iteration  |      lk      |    lk-lko    |
 |--------------|--------------|--------------|
 |            1 |     -740.781 |          Inf |
 |            2 |     -732.495 |      8.28629 |
 |            3 |      -732.49 |   0.00541045 |
 |            4 |      -732.49 |   9.8679e-09 |
 |--------------|--------------|--------------|
Second step estimation
 |--------------|--------------|--------------|
 |   iteration  |      lk      |    lk-lko    |
 |--------------|--------------|--------------|
 |            1 |     -552.702 |          Inf |
```

```
 |            2 |     -528.266 |      24.4361 |
 |            3 |     -513.702 |      14.5641 |
 |            4 |     -509.195 |      4.50721 |
 |            5 |     -509.192 |   0.00285414 |
 |            6 |     -509.192 |  1.11389e-08 |
 |--------------|--------------|--------------|
```

The first panel reports the iterations of the first step CML estimation of the regression coefficients in the static logit model, while the second refers to the second step maximization to obtain the pseudo-CML estimates of the parameters in Equation 10.

After calling `summary(out5)`, the following results are displayed:

```
Call:
cquad_pseudo(id = id, yv = yv, X = X)

Log-likelihood:
-509.1917
                  est.      s.e.      t-stat      p-value
married     0.19259731 0.1858896  1.0360844 3.001628e-01
year21982   0.05031661 0.2664274  0.1888567 8.502051e-01
year21983 -0.12381494 0.2092980 -0.5915724 5.541369e-01
year21984 -0.02956563 0.2224643 -0.1329006 8.942720e-01
year21985 -0.43257573 0.2243302 -1.9282989 5.381796e-02
year21986 -0.54727988 0.2212247 -2.4738647 1.336603e-02
year21987  0.17223711 0.2425840  0.7100103 4.776978e-01
y_lag      1.47526322 0.1807924  8.1599843 4.440892e-16
```

Notice that the estimation results are in agreement with those obtained by fitting the QEext or the QEbasic models; however they exhibit some differences since the pseudo-CML estimator is based on the conditional probability in Equation 11 that contains the parameters of the true DL model. Nevertheless, these results confirm the presence of a high degree of state dependence in union participation.

4.2. Use of `sim_panel_logit` to generate dynamic binary panel data

In the following, we illustrate how to perform a simple simulation study on data generated from a DL model by means of function `sim_panel_logit` in package **cquad**. In this example, we fit the modified QEequ model by CML and study the properties of the test for state dependence proposed by Bartolucci *et al.* (2017). The script to replicate the exercise is reported below

```
R> require(cquad)
R> n <- 500
R> TT <- 6
R> nit <- 100
R> be <- 1
R> rho <- 0.5
```

```
R> var <- (pi * pi) / 3
R> stdep <- c(0, 1)
R> TEST <- rep(0, nit)
R> for (ga in stdep) {
+    for (it in 1:nit) {
+      label <- 1:n
+      id <- rep(label, each = TT)
+      X <- matrix(rep(0), n * TT, 1)
+      alpha <- rep(0, n)
+      eta <- rep(0, n * TT)
+      e <- rnorm(n * TT) * sqrt(var * (1 - rho^2))
+      j <- 0
+      for (i in 1:n) {
+        j <- j + 1
+        X[j] <- rnorm(1) * sqrt(var)
+        for (t in 2:TT) {
+          j <- j + 1
+          X[j] <- rho * X[j - 1] + e[j]
+        }
+        alpha[i] <- (X[j - 2] + X[j - 1] + X[j]) / 3
+      }
+      cat("sample n. ", it, "\n")
+      data <- sim_panel_logit(id, alpha, X, c(be, ga), dyn = TRUE)
+      yv <- data$yv
+      mod <- cquad(yv ~ X, data.frame(yv, X), index = 500, model = "equal")
+      beta <- mod$coefficients
+      TEST[it] <- beta[length(beta)]/mod$se[length(beta)]
+    }
+   cat(c("gamma =", ga, "\n"))
+   RES <- c(mean(TEST), mean(abs(TEST) > 1.96))
+   names(RES) <- c("t-stat", "rej. rate")
+   print(RES)
+ }
```

In the first part of the script, we set the simulation parameters for the sample size, number of time occasions and number of Monte Carlo replications. We also set the parameter values for the DL model in Equation 2 with one regression parameter $\beta = 1$ and one covariate, generated as an AR(1) process with autocorrelation coefficient $\rho = 0.5$. In this exercise, we analyze two scenarios, with the state dependence parameter γ equal to 0 and 1.

In the first part of the script inside the `for` loops, we generate the identifier `id` as an n-dimensional vector, the $n \times T$ vector for the single covariate `X`, and the n-dimensional vector of individual intercepts `alpha`, which is computed in a similar manner as in Honoré and Kyriazidou (2000). Lastly, we generate the binary response variable using function `sim_panel_logit` described in Section 3.1. As the function returns both the binary variable and the response probabilities, the dependent variable needs to be retrieved by `yv <- data$yv`.

Once the data have been generated, we proceed to the estimation of the QEequ model using `cquad` with `model = "equal"` to fit the modified QE model in Equation 7 by CML; we store the results for the t test in Equation 9. Finally, we display the results containing the average value of the test in the 100 sample and the average rejection rate of a bilateral test at the 0.05 significance level. The last part or the script produces the following output:

```
...

gamma = 0
    t-stat  rej. rate
-0.1753164  0.0400000

...

gamma = 1
   t-stat rej. rate
 4.939813  0.990000
```

where the iteration logs from `cquad` have been omitted. Under the null hypothesis $\gamma = 0$, the rejection rate is very close to the nominal size of 0.05, while under the alternative hypothesis $\gamma = 1$ the test exhibits good power properties. These results are close to those found by Bartolucci *et al.* (2017) in their simulation study, to which we refer the reader for an extension of this simple design to several other scenarios.

4.3. Analysis of `union` data in Stata

In the following, we illustrate the Stata module **cquad** that contains the four commands to fit the QE models described in Section 2.3 by an example based again on data about unionized workers. The dataset to replicate this example is already available in the Stata online data repository and is contained in file `union.dta`.

The three commands reported below load the dataset, then describe the panel structure, already in place, and list the variables present in the dataset

```
webuse union
xtdes
descr
```

The output generated by these command lines is:

```
. webuse union
(NLS Women 14-24 in 1968)

. xtdes

  idcode:  1, 2, ..., 5159                                    n =       4434
    year:  70, 71, ..., 88                                    T =         12
           Delta(year) = 1 unit
           Span(year)  = 19 periods
           (idcode*year uniquely identifies each observation)
```

```
Distribution of T_i:   min      5%     25%      50%      75%      95%      max
                         1       1       3        6        8       11       12

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------------------
      190      4.29    4.29 |  1111...11.1.11.1.11
      129      2.91    7.19 |  .......11.1.11.1.11
       93      2.10    9.29 |  1..................
       78      1.76   11.05 |  .......1..........
       68      1.53   12.58 |  ..11...11.1.11.1.11
       64      1.44   14.03 |  ...1...11.1.11.1.11
       60      1.35   15.38 |  .111...11.1.11.1.11
       52      1.17   16.55 |  11.................
       52      1.17   17.73 |  1111...............
     3648     82.27  100.00 | (other patterns)
 ---------------------------+---------------------
     4434    100.00         |  XXXX...XX.X.XX.X.XX

. descr

Contains data from http://www.stata-press.com/data/r13/union.dta
  obs:        26,200                          NLS Women 14-24 in 1968
 vars:             8                          4 May 2013 13:54
 size:       235,800
--------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------
idcode          int     %8.0g                 NLS ID
year            byte    %8.0g                 interview year
age             byte    %8.0g                 age in current year
grade           byte    %8.0g                 current grade completed
not_smsa        byte    %8.0g                 1 if not SMSA
south           byte    %8.0g                 1 if south
union           byte    %8.0g                 1 if union
black           byte    %8.0g                 race black
--------------------------------------------------------------------------------
Sorted by:  idcode  year
```

The dataset consists of 4434 women between 14 and 24 years old in 1968, interviewed between 1970 and 1988. The panel is unbalanced and the maximum number of occasions of observation of the same subject is 12. The last part of the output reports the variable description, where `union` is the response variable in our exercise, `age`, `grade`, `not_smsa`, and `south` are the covariates, while `black` is excluded from the analysis because of its time-invariant nature.

We first illustrate command `cquadbasic` to fit the QEbasic model in Equation 6 by CML, where we include time dummies in the model specification by using the `xi` and `i.year` declarations. The command line *xi: cquadbasic union idcode age grade south not_smsa i.year*

produces the following output

```
. xi: cquadbasic union idcode age grade south not_smsa i.year
i.year            _Iyear_70-88       (naturally coded; _Iyear_70 omitted)

Fit (simplified) quadratic exponential model by Conditional Maximum Likelihood
see Bartolucci & Nigro (2010), Econometrica

      |      lk       lk-lk0
------+----------------------------------------------------
    1 | -3439.9096  1.000e+10
    2 | -3071.6412  368.26839
    3 | -3069.0539  2.5872579
    4 | -3069.0534  .00050444
    5 | -3069.0534  5.775e-11


             |      est.      s.e       t-stat.    p-value
-------------+----------------------------------------------------
         age |  .17670917   .1192216  1.4821908  .06914476
       grade | -.03658997  .04586492 -.79777692  .21249998
       south |  -.5191613  .13732314 -3.7805814  .00007823
    not_smsa |  .12631127  .13146408   .9608044  .16832526
   _Iyear_71 |  1.5208636   1.035464  1.4687749  .07094693
   _Iyear_72 |  1.1096837  .91812295  1.2086439  .11339984
   _Iyear_73 |  .90256541  .79733234  1.1319814  .12882112
   _Iyear_77 |   .1829554   .3308496  .55298662  .29013629
   _Iyear_78 |  .17904624  .21288676   .8410398  .20016282
   _Iyear_80 |  .46950024   .0846961  5.5433514  1.484e-08
   _Iyear_82 | -.40988205  .28664089 -1.4299497  .07636573
   _Iyear_83 | -.95438994  .40301052 -2.3681514  .00893861
   _Iyear_85 | -.73765258  .63762525 -1.1568748  .12366176
   _Iyear_87 | -1.3247366  .87641421 -1.5115416  .06532525
   _Iyear_88 | -.93795347  1.0381108 -.90351959   .1831251
       y-lag |  1.5332567  .06307817  24.307248          0
```

First the iteration logs are reported, then the estimation output is displayed in a standard fashion, reporting the estimated coefficients for the QEbasic model, along with asymptotic standard errors, the related t test statistics and p values. Notice that the estimate associated with ψ in Equation 6 reflects a high degree of positive state dependence, in line with the well-known results in other applied works.

The extended version of the QE model, QEext, can be fitted in a similar manner, by using command `cquadext`:

cquadext union idcode age grade south not_smsa _Iyear_72 _Iyear_73 _Iyear_77 _Iyear_78 _Iyear_80 _Iyear_82 _Iyear_83 _Iyear_85 _Iyear_87

Notice that here we are not using the `xi:` prefix and the factor `i.year` as explanatory variable. In fact, we list the time dummies separately in order to exclude the dummy for 1988: in the QEext model, not all the effects associated with the time dummies can be identified, due to the presence of an intercept term, ϕ, in the regressors referred to the observation at time T (see Equation 4).

The above code produces the following output:

```
. cquadext union idcode age grade south not_smsa _Iyear_72 _Iyear_73
> 2 _Iyear_77 _Iyear_78 _Iyear_80 _Iyear_8 _Iyear_83 _Iyear_85 _Iyear_87

Fit quadratic exponential model by Conditional Maximum Likelihood
see Bartolucci & Nigro (2010), Econometrica
```

(output omitted)

```
               |      est.      s.e.      t-stat.    p-value
---------------+-----------------------------------------------
           age |  .17308473  .11933765  1.4503782  .07347655
         grade | -.04047509   .0465145 -.87016079  .19210627
         south | -.51184847  .13953697 -3.6681926  .00012214
      not_smsa |  .17524652  .13523937  1.2958248  .09751793
     _Iyear_72 |  -.4644361   .1964388 -2.3642789   .0090326
     _Iyear_73 | -.65950516  .27895047 -2.3642375  .00903361
     _Iyear_77 | -1.3784265  .72358421 -1.9049981  .02839016
     _Iyear_78 | -1.3701126  .84614133 -1.6192479  .05269697
     _Iyear_80 | -1.1167485  1.0780889 -1.0358595  .15013386
     _Iyear_82 | -1.9383478  1.3150617 -1.4739595  .07024624
     _Iyear_83 | -2.4862166   1.433189 -1.7347444  .04139305
     _Iyear_85 |  -2.293721  1.6709237 -1.3727264  .08491871
     _Iyear_87 | -2.8867738  1.9100228 -1.5113819  .06534559
      diff-int | -2.9745408  2.2316307 -1.3329001   .0912823
      diff-age |  .01050808  .02053247  .51177844  .30440304
    diff-grade |  .01403913  .02483142  .56537754   .2859085
    diff-south | -.01017179  .12702618 -.08007635  .46808827
 diff-not_smsa | -.24502608  .14435482 -1.6973876   .0448117
diff-_Iyear_72 |  4.5353507  2.3909244  1.8969026   .0289204
diff-_Iyear_73 |   3.213293  2.1675763  1.4824359  .06911217
diff-_Iyear_77 |  2.9858792  2.1187489  1.4092653  .07937837
diff-_Iyear_78 |  2.8557536  2.1441469  1.3318834  .09144926
diff-_Iyear_80 |  3.4787453  2.1165322  1.6436061   .0501288
diff-_Iyear_82 |  2.3113123   2.108686  1.0960913  .13651941
diff-_Iyear_83 |  2.4132524  2.1023472  1.1478848  .12550806
diff-_Iyear_85 |  2.7441521  2.0885294   1.313916  .09443724
diff-_Iyear_87 |  2.6838554   2.079152  1.2908414  .09837934
         y-lag |  1.5646588  .06439017  24.299654          0
```

where the iteration logs have been omitted for brevity. If the time-dummy associated with the last observation is not dropped beforehand, a warning message is printed, and the results are obtained using the generalized inverse of the Hessian.

The modified QE model, QEequ, can be estimated by calling `cquadequ`:

`xi: cquadequ union idcode age grade south not_smsa i.year`

```
. xi: cquadequ union idcode age grade south not_smsa i.year
i.year            _Iyear_70-88        (naturally coded; _Iyear_70 omitted)
```

(output omitted)

```
             |      est.      s.e       t-stat.    p-value
-------------+---------------------------------------------------
         age |  .16845566  .11901965  1.4153601  .07848147
       grade | -.03958659  .04550678 -.86990548  .19217603
       south | -.53406297  .13625918  -3.919464  .00004437
    not_smsa |   .0984639  .13080979  .75272577  .22580736
   _Iyear_71 |  1.6032853  1.0337023  1.5510126  .06044933
   _Iyear_72 |  1.1740137  .91650676  1.2809657  .10010286
   _Iyear_73 |  .97015581  .79589985  1.2189421  .11143309
   _Iyear_77 |  .24177005  .33043231  .73167798  .23218257
   _Iyear_78 |  .25282926  .21264697  1.1889624  .11722723
   _Iyear_80 |  .54363568  .08483378  6.4082453  7.360e-11
   _Iyear_82 |  -.3246461   .2861711 -1.1344475  .12830343
   _Iyear_83 | -.88650878  .40228033  -2.203709  .01377241
   _Iyear_85 | -.68779397  .63653421 -1.0805295  .13995324
   _Iyear_87 | -1.3316314  .87497451 -1.5219087  .06401597
   _Iyear_88 | -1.5551096  1.0362781 -1.5006681  .06672071
       y-lag |  .76891417  .03180295  24.177448          0
```

The estimation results are different from those obtained by `cquadbasic` because of the different way the association between y_{it} and y_{it-1} is specified in Equation 7. The test for absence of state dependence is the t test associated with the lagged dependent variable reported in the output above.

Finally, command `cquadpseudo` fits the pseudo-CML estimator of the parameters of the DL model described in Section 2.4. The input line is as follows

`xi: cquadpseudo union idcode age grade south not_smsa i.year`

and produces the following output:

```
. xi: cquadpseudo union idcode age grade south not_smsa i.year
i.year            _Iyear_70-88        (naturally coded; _Iyear_70 omitted)
```

```
Fit Pseudo Conditional Maximum Likelihood estimator for the dynamic logit model
see Bartolucci & Nigro (2012), J.Econometrics

First step

      |      lk       lk-lk0
------+-------------------------------------------------
    1 | -4550.1859  1.000e+10
    2 | -4508.4587  41.727174
    3 | -4479.6267  28.832058
    4 | -4464.3395  15.287228
    5 | -4462.0772  2.2622144
    6 |  -4462.077   .0002431
    7 |  -4462.077  1.000e-11

Second step

      |      lk       lk-lk0
------+-------------------------------------------------
    1 | -3386.3831  1.000e+10
    2 | -3072.2352  314.14795
    3 | -3068.2783  3.9568752
    4 | -3068.2768  .00144833
    5 | -3068.2768  5.689e-10


             |      est.     s.e.(rob)    t-stat.    p-value
-------------+-------------------------------------------------
         age |  .18590097  .12502643  1.4868934  .13704297
       grade | -.03115066  .05488738 -.56753782  .57034884
       south | -.62116171  .16083689 -3.8620598  .00011244
    not_smsa |  .10764683  .14923884  .72130574  .47072142
   _Iyear_71 |  .66895192  1.0824925  .61797374  .53659265
   _Iyear_72 |  .26741545  .96467342  .27720827  .78162019
   _Iyear_73 |  .04473093  .83125482  .05381134  .95708548
   _Iyear_77 | -.66439033   .3474518 -1.9121798  .05585313
   _Iyear_78 | -.56283602  .22525051 -2.4987114  .01246458
   _Iyear_80 | -.42448135  .08815153 -4.8153602  1.469e-06
   _Iyear_82 | -1.3962766  .30058041 -4.6452681  3.396e-06
   _Iyear_83 | -1.8777382  .42388142 -4.4298667  9.429e-06
   _Iyear_85 | -1.7545693     .66529    -2.6373  .00835689
   _Iyear_87 |  -2.409943  .91783499 -2.6256822  .00864755
   _Iyear_88 | -2.5102739  1.0890873 -2.3049337  .02117029
       y-lag |  1.6295114  .07720721  21.105691          0
```

The first part of the output reports the value of the log-likelihood at each iteration for the first step, the CML estimation of the regression coefficients using a static logit model, while the second refers to the maximization of the pseudo log-likelihood with respect to the parameters in Equation 10. The estimation results are similar to those obtained with the QE model.

5. Acknowledgments

We acknowledge the financial support from the grant RBFR12SHVV of the Italian Government (FIRB project "Mixture and latent variable models for causal inference and analysis of socio-economic data").

References

Andersen EB (1970). "Asymptotic Properties of Conditional Maximum-Likelihood Estimators." *Journal of the Royal Statistical Society B*, **32**(2), 283–301.

Barndorff-Nielsen O (1978). *Information and Exponential Families in Statistical Theory*. John Wiley & Sons. doi:10.1002/9781118857281.

Bartolucci F (2015). "**cquad**: Stata Module to Perform Conditional Maximum Likelihood Estimation of Quadratic Exponential Models." Statistical Software Components, Boston College Department of Economics. URL https://ideas.repec.org/c/boc/bocode/s457891.html.

Bartolucci F, Bellio R, Salvan A, Sartori N (2016). "Modified Profile Likelihood for Fixed-Effects Panel Data Models." *Econometric Reviews*, **35**(7), 1271–1289. doi:10.1080/07474938.2014.975642.

Bartolucci F, Nigro V (2010). "A Dynamic Model for Binary Panel Data with Unobserved Heterogeneity Admitting a $\sqrt{n}$-Consistent Conditional Estimator." *Econometrica*, **78**(2), 719–733. doi:10.3982/ecta7531.

Bartolucci F, Nigro V (2012). "Pseudo Conditional Maximum Likelihood Estimation of the Dynamic Logit Model for Binary Panel Data." *Journal of Econometrics*, **170**(1), 102–116. doi:10.1016/j.jeconom.2012.03.004.

Bartolucci F, Nigro V, Pigini C (2017). "Testing for State Dependence in Binary Panel Data with Individual Covariates." *Econometric Reviews*. doi:10.1080/07474938.2015.1060039. Forthcoming.

Bartolucci F, Pigini C (2017). ***cquad**: Conditional Maximum Likelihood for Quadratic Exponential Models for Binary Panel Data*. R package version 1.4, URL https://CRAN.R-project.org/package=cquad.

Chamberlain G (1980). "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies*, **47**(1), 225–238. doi:10.2307/2297110.

Chamberlain G (1985). "Heterogeneity, Omitted Variable Bias, and Duration Dependence." In JJ Heckman, BS Singer (eds.), *Longitudinal Analysis of Labor Market Data*, Econometric Society Monographs, pp. 3–38. Cambridge University Press, Cambridge. doi:10.1017/ccol0521304539.001.

Cox DR (1972). "The Analysis of Multivariate Binary Data." *Journal of the Royal Statistical Society C*, **21**(2), 113–120. doi:10.2307/2346482.

Croissant Y, Millo G (2008). "Panel Data Econometrics in R: The **plm** Package." *Journal of Statistical Software*, **27**(2), 1–43. doi:10.18637/jss.v027.i02.

Heckman JJ (1981a). "Heterogeneity and State Dependence." In S Rosen (ed.), *Studies in Labor Markets*, pp. 91–140. University of Chicago Press. URL http://www.nber.org/chapters/c8909.

Heckman JJ (1981b). "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process." In CF Manski, D McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 179–195. MIT Press, Cambridge.

Honoré BE, Kyriazidou E (2000). "Panel Data Discrete Choice Models with Lagged Dependent Variables." *Econometrica*, **68**(4), 839–874. doi:10.1111/1468-0262.00139.

Hsiao C (2005). *Analysis of Panel Data*. 2nd edition. Cambridge University Press, New York.

Lucchetti R, Pigini C (2015). "**DPB**: Dynamic Panel Binary Data Models in **gretl**." ***gretl** working paper 1*, Università Politecnica delle Marche (I), Dipartimento di Scienze Economiche e Sociali. URL https://ideas.repec.org/p/anc/wgretl/1.html.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2017). ***nlme**: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-131, URL https://CRAN.R-project.org/package=nlme.

Rabe-Hesketh S (2011). "**GLLAMM**: Stata Program to Fit Generalised Linear Latent and Mixed Models." Statistical Software Components, Boston College Department of Economics. URL https://ideas.repec.org/c/boc/bocode/s401701.html.

Rabe-Hesketh S, Skrondal A, Pickles A (2005). "Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects." *Journal of Econometrics*, **128**(2), 301–323. doi:10.1016/j.jeconom.2004.08.017.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

StataCorp (2015). *Stata Statistical Software: Release 14*. StataCorp LP, College Station, TX. URL http://www.stata.com/.

Stewart M (2006). "Maximum Simulated Likelihood Estimation of Random-Effects Dynamic Probit Models with Autocorrelated Errors." *Stata Journal*, **6**(2), 256–272.

White H (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, **48**(4), 817–838. doi:10.2307/1912934.

Wooldridge JM (2005). "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics*, **20**(1), 39–54. doi:10.1002/jae.770.

Affiliation:

Francesco Bartolucci
Department of Economics
University of Perugia
06123 Perugia, Italia
E-mail: francesco.bartolucci@unipg.it
URL: https://sites.google.com/site/bartstatistics/

Claudia Pigini
Department of Economics and Social Sciences
Marche Polytechnic University
60121 Ancona, Italia
E-mail: c.pigini@univpm.it
URL: http://www.univpm.it/claudia.pigini

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
June 2017, Volume 78, Issue 7 *Submitted:* 2015-07-27
doi:10.18637/jss.v078.i07 *Accepted:* 2016-04-01

Granger causality in dynamic binary short panel data models

Francesco Bartolucci and Claudia Pigini

University of Perugia, Marche Polytechnic University

13 March 2017

Granger causality in dynamic binary short panel data models

Francesco Bartolucci
Università di Perugia (IT)
francesco.bartolucci@unipg.it

Claudia Pigini
Università Politecnica
delle Marche and MoFiR (IT)
c.pigini@univpm.it

March 13, 2017

Abstract

Strict exogeneity of covariates other than the lagged dependent variable, and conditional on unobserved heterogeneity, is often required for consistent estimation of binary panel data models. This assumption is likely to be violated in practice because of feedback effects from the past of the outcome variable on the present value of covariates and no general solution is yet available. In this paper, we provide the conditions for a logit model formulation that takes into account feedback effects without specifying a joint parametric model for the outcome and predetermined explanatory variables. Our formulation is based on the equivalence between Granger's definition of noncausality and a modification of the Sims' strict exogeneity assumption for nonlinear panel data models, introduced by Chamberlain (1982) and for which we provide a more general theorem. We further propose estimating the model parameters with a recent fixed-effects approach based on pseudo conditional inference, adapted to the present case, thereby taking care of the correlation between individual permanent unobserved heterogeneity and the model's covariates as well. Our results hold for short panels with a large number of cross-section units, a case of great interest in microeconomic applications.

Keywords: fixed effects, noncausality, predetermined covariates, pseudo-conditional inference, strict exogeneity.

JEL Classification: C12, C23, C25

1 Introduction

There is an increasing number of empirical microeconomic applications that require the estimation of binary panel data models, which are typically dynamic so as to account for state dependence (Heckman, 1981).[1] In these contexts, strict exogeneity of covariates other than the lagged dependent variable, conditional on unobserved heterogeneity, is required for consistent estimation of the regression and state dependence parameters, when the estimation relies on correlated random effects or on fixed effects which are eliminated when conditioning on suitable sufficient statistics for the individual unobserved heterogeneity. However, the assumption of strict exogeneity is likely to be violated in practice because there may be feedback effects from the past of the outcome variable on the present values of the covariates, namely the model covariates may be Granger-caused by the response variable Granger (1969). While in linear models the mainstream approach to overcome this problem is to consider instrumental variables (Anderson and Hsiao, 1981; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998), considerably fewer results are available for nonlinear binary panel data models with predetermined covariates. This is particularly true with short binary panel data and no general solution is yet available, despite the relevance of binary these type of data in microeconomic applications.

Honoré and Lewbel (2002) propose a semiparametric estimator for the parameters of a binary choice model with predetermined covariates. However, they provide 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) develop a semiparametric strategy based on the 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 when the set of relevant explanatory variables is large. A different approach is taken by Wooldridge (2000), who proposes 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), to account for the dependence between strictly exogenous explanatory variables and individual unobserved effects, combined with a preliminary version of the Wooldridge (2005)'s

[1] Estimators of dynamic discrete choice models are employed in studies related to labor market participation (Heckman and Borjas, 1980; Arulampalam, 2002; Stewart, 2007), and specifically to female labor supply and fertility choices (Hyslop, 1999; Carrasco, 2001; Keane and Sauer, 2009; Michaud and Tatsiramos, 2011), self-reported health status (Contoyannis et al., 2004; Halliday, 2008; Heiss, 2011; Carro and Traferri, 2012), poverty traps (Cappellari and Jenkins, 2004; Biewen, 2009), welfare participation (Wunder and Riphahn, 2014), unionization of workers (Wooldridge, 2005), household finance (Alessie et al., 2004; Giarda, 2013; Brown et al., 2014), firms' access to credit (Pigini et al., 2016), and migrants' remitting behavior (Bettin and Lucchetti, 2016)

solution to the initial conditions problem. Although this is an intuitive strategy, it relies on distributional assumptions on the individual unobserved heterogeneity; moreover, it is computationally demanding when the number of predetermined covariates is large and it requires strict exogeneity of the covariates used for the parametric random-effects correction.

A strategy similar to that developed by Wooldridge (2000) is adopted by Mosconi and Seri (2006), who test for the presence of feedback effects in binary bivariate time-series by means of Maximum Likelihood (ML)-based test statistics. They build their estimation and testing proposals 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 intertemporal choices related to the labor market, poverty traps, and persistence in unemployment, often rely on very short time-series and a large number of cross-section units resulting from rotated surveys. Furthermore, in the short panel data setting, 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. Moreover, 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 dynamic binary fixed T-panel data model that takes into account general forms of feedback effects from the past of the outcome variable on the present value of the covariates. Our formulation presents three main advantages with respect to the available solutions. First, it does not require the specification of a joint parametric model for the outcome and predetermined explanatory variables. In fact, the starting point to build the proposed formulation is the definition of noncausality (Granger, 1969), the violation of which corresponds to the presence of feedback effects, as stated in terms of conditional independence by Chamberlain (1982) for nonlinear models. Translating the definition of noncausality to a parametric model requires, however, the specification of the conditional probability for the covariates (x). On the contrary, we follow Chamberlain (1982) and introduce an equivalent definition based on a modification of Sims (1972)'s strict exogeneity for nonlinear models, which only involves specifying the probability for the binary dependent variable at each time occasion (y_t) conditional on past, present, and future values of x, and for which we provide a more general theorem of equivalence to noncausality.

Second, the proposed model has a simple formulation and allows for the inclusion of even a large number of predetermined covariates. Under the logit model, it amounts to augment the linear index function with a linear combination of the leads of the predetermined covariates, along with the lags of the binary dependent variable. We analytically prove that this augmented linear index function corresponds to the logit for the joint distribution of y_t and the 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) and that their conditional means depend on time-fixed effects. In the other cases, we anyway assume a linear approximation which proves to be effective in series of simulations while allowing us to maintain a simple approach.

Third, the logit formulation allows for a fixed-effects estimation approach based on sufficient statistics for the incidental parameters, thus avoiding parametric assumptions on the distribution of the individual unobserved heterogeneity. In particular, we propose estimating the model parameters by means of a Pseudo Conditional Maximum Likelihood (PCML) estimator recently put forward by Bartolucci and Nigro (2012), and here adapted to the proposed extended formulation. They approximate the dynamic logit with a Quadratic Exponential (QE) model (Cox, 1972; Bartolucci and Nigro, 2010), which admits a sufficient statistics for the incidental parameters and has the same interpretation as the dynamic logit model in terms of log-odds ratio between pairs of consecutive outcomes. In simpler contexts, this approach leads to a consistent estimator of the model parameters under the null hypothesis of absence of true state dependence, whereas has a reduced bias even with strong state dependence.

We study the finite sample properties of the PCML estimator for the proposed model through an extensive simulation study. The results show that the PCML estimator exhibits a negligible bias, for both the regression parameter associated with the predetermined covariate and the state dependence parameter, in the presence of substantial departures from noncausality. In addition, the estimation bias is almost negligible when the density of the predetermined covariate does not belong to the exponential family or its conditional mean depends on time-varying effects. It is also worth noting that the qualities of the proposed approach emerge for quite short T and a large number of cross-section units. Finally, the PCML is compared with the correlated random-effects ML estimator of Wooldridge (2005), adapted for the proposed formulation. This ML estimator is consistent for the parameters of interest in presence of feedbacks, although remarkably less efficient than the PCML in estimating the state dependence parameter, especially with short T. However, differently from our approach, consistency relies on the assumption of independence between the predetermined covariates and the individual unobserved effects, which is hardly tenable in practice.

The rest of the paper is organized as follows. Section 2 introduces the definitions of noncausality and strict exogeneity for nonlinear models. In Section 3 we illustrate the proposed model formulation. Section 4 describes the PCML estimation approach. Section

5 outlines the simulation study, and Section 6 provides main conclusions.

2 Definitions

Consider panel data for a sample of n units observed at T occasions according to 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 the same sample unit i. Note that here we consider only one covariate to maintain 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 in Chamberlain (1982), assuming that the economic life of any individual begins at time $t = 1$, the Granger's definition of noncausality is:

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:*

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

Testing for G requires the knowledge and formulation of the model for each time-specific covariate given the the previous covariates and responses. However, following Chamberlain (1982), we introduce a condition that is the basis of the approach that we present in the next sections.

Definition. S' - *x is strictly exogenous 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*

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

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

The following result holds, whose proof is related to that provided in Chamberlain (1982).

Theorem 1. G *and* S' *are equivalent conditions.*

Proof. G may be reformulated as

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

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 (1) and then G. □

It is worth noting that, apart from the case $T = 2$, definition S' is stronger than the definition of strict exogeneity of Sims (1972) adapted to the case of binary panel data, which we denote by 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:

Definition. S - *x is strictly exogenous 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*

$$p(y_{it}|c_i, \boldsymbol{x}_i) = p(y_{it}|c_i, \boldsymbol{x}_{i,1:t}), \quad i = 1, \ldots, n,\ t = 2, \ldots, T. \tag{3}$$

Theorem 2. G *implies* S.

Proof. Proceeding as in the proof of Theorem 1, G implies that

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

By recursively using the previous expression for a fixed s and for t from $T-1$ to s, we obtain condition (3). □

Although the focus here is on nonlinear binary panel data models, it is useful to accompany the discussion with the Granger's and the Sims' definitions in the simpler context of linear models, as laid out by Chamberlain (1984), where testable restrictions on the regression parameters can be derived directly. The starting point is a linear panel data model of the form

$$y_{it} = x_{it}\beta + c_i + \varepsilon_{it}, \quad i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{4}$$

where now the dependent variables y_{it} are continuous and the error terms ε_{it} are iid. The usual exogeneity assumption is stated as

$$\mathrm{E}(\varepsilon_{it}|c_i, \boldsymbol{x}_i) = 0, \quad i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{5}$$

which rules out the lagged response variables from the regression specification, as well as possible feedback effects from past values of y_{it} on to the present and future values of the covariate.

Now consider the minimum mean-square error linear predictor, denoted by $\mathrm{E}^*(\cdot)$, and consider the following definitions, which hold for all i:

$$\mathrm{E}^*(c_i|\boldsymbol{x}_i) = \eta + \boldsymbol{x}_i'\boldsymbol{\lambda}, \tag{6}$$

$$\mathrm{E}^*(y_{it}|\boldsymbol{x}_i) = \alpha_t + \boldsymbol{x}_i'\boldsymbol{\pi}_t, \quad t = 1, \ldots, T, \tag{7}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ and $\boldsymbol{\pi}_t = (\pi_{t1}, \ldots, \pi_{tT})'$ are vectors of regression coefficients. Equation (7) may also be expressed as

$$\mathrm{E}^*(\boldsymbol{y}_i|\boldsymbol{x}_i) = \boldsymbol{\alpha} + \boldsymbol{\Pi}\boldsymbol{x}_i,$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)'$ and $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_T)'$. It may be simply proved that assumptions (4), (5), together with definition (6), imply that

$$\boldsymbol{\Pi} = \beta\boldsymbol{I} + \boldsymbol{1}\boldsymbol{\lambda}',$$

where $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{1}$ is a column vector of ones of suitable dimension; in the present case they are of dimension T. In Chamberlain (1984), the structure of $\boldsymbol{\Pi}$ is related to the definition of strict exogeneity in Sims (1972) for linear models (equivalent to condition S for binary models defined above) that, conditional on the permanent unobserved heterogeneity, is stated as

$$\mathrm{E}^*\left(y_{it}|c_i, \boldsymbol{x}_i\right) = \mathrm{E}^*\left(y_{it}|c_i, \boldsymbol{x}_{i,1:t}\right), \quad t = 1, \ldots, T. \tag{8}$$

Sims (1972) proved the equivalence of this condition with that of noncausality of Granger

(1969). In matrix notation, condition (8) can be written as

$$\mathrm{E}^*(\boldsymbol{y}_i|c_i, \boldsymbol{x}_i) = \boldsymbol{\varphi} + \boldsymbol{\Psi}\boldsymbol{x}_i + c_i\boldsymbol{\tau}, \tag{9}$$

where $\boldsymbol{\Psi}$ is a lower triangular matrix, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_T)'$, and $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_T)'$. Assumptions (6) and (9) then imply the following structure for $\boldsymbol{\Pi}$:

$$\boldsymbol{\Pi} = \boldsymbol{B} + \boldsymbol{\delta}\boldsymbol{\lambda}',$$

where $\boldsymbol{B}$ is a lower triangular matrix and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_T)'$.

It is straightforward to translate the restrictions in the structure of $\boldsymbol{\Pi}$ to the linear index function of a nonlinear model. In fact, Chamberlain (1984) and then Wooldridge (2010, Section 15.8.2) show that a simple test for strict exogeneity, S, in binary panel data models can be readily derived by adding $\boldsymbol{x}_{i,t+1}$ to the set of explanatory variables. In the next section we show not only that noncausality S' can be tested in a similar manner within a dynamic model formulation, but also that the linear index augmented with $\boldsymbol{x}_{i,t+1}$ represents, under rather general conditions, the exact log-odds ratio for the joint probability of y_{it} and $\boldsymbol{x}_{i,t+1}$ when S' is violated, thereby providing a model formulation that accounts for feedback effects and whose parameters may be consistently estimated.

3 Model formulation

Consider the general case in which, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, 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 $\boldsymbol{X}_{i,t_1:t_2} = (\boldsymbol{x}_{it_1}, \ldots, \boldsymbol{x}_{it_2})$, with $\boldsymbol{X}_i = \boldsymbol{X}_{i,1:T}$ being the matrix of the covariates for all time occasions. In order to illustrate the proposed model, we first recall the main assumptions of the dynamic logit model.

3.1 Dynamic logit model

A standard formulation of a dynamic binary choice model assumes that, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, the binary response y_{it} has conditional distribution

$$p(y_{it}|c_i, \boldsymbol{X}_i, \boldsymbol{y}_{i,1:t-1}) = p(y_{it}|c_i, \boldsymbol{x}_{it}, y_{i,t-1}), \tag{10}$$

corresponding to a first-order Markov model for y_{it} with dependence only on the present values of the explanatory variables. The above conditioning set can be easily enlarged to include further lags of $\boldsymbol{x}_{it}$ and y_{it}.

Moreover, adopting a logit formulation for the conditional probability (see Hsiao, 2005, ch. 7, for a review), that is,

$$p(y_{it}|c_i, \boldsymbol{x}_{it}, y_{i,t-1}) = \frac{\exp\left[y_{it}\left(c_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma\right)\right]}{1 + \exp\left(c_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma\right)}, \quad t = 2, \ldots, T, \tag{11}$$

the conditional distribution of the overall vector of responses becomes:

$$p(\boldsymbol{y}_{i,2:T}|c_i, \boldsymbol{X}_i, y_{i1}) = \frac{\exp\left[y_{i+}c_i + \sum_{t=2}^{T} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{it-1}\gamma\right)\right]}{\prod_{t=2}^{T}\left[1 + \exp\left(c_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma\right)\right]}, \tag{12}$$

where $\boldsymbol{\beta}$ and γ are the parameters of interest for the covariates and the true state dependence (Heckman, 1981), respectively, $y_{i+} = \sum_{t=2}^{T} y_{it}$ is the *total score* and the individual-specific intercepts c_i are often considered as nuisance parameters; moreover, the initial observation y_{i1} is considered as given.

Expression (10) embeds assumption S' by excluding leads of $\boldsymbol{x}_{it}$ from the probability conditioning set. It therefore rules out feedbacks from the response variable to future covariates, that is, the Granger causality. Noncausality is often a hardly tenable assumption, as when the covariates of interest depend on individual choices. If covariates are predetermined, as opposed to strictly exogenous, estimation of the model parameters of interest can be severely biased, when estimation is based on eliminating or approximating c_i with quantities depending on the entire observed history of covariates (Mundlak, 1978; Chamberlain, 1984; Wooldridge, 2005).

3.2 Proposed model

As stated at the end of Section 2, dealing with violations of condition S', formulated as in (2), amounts to propose a generalization of the standard dynamic binary choice model based on assumption (10). In order to allow for such violations, we specify the probability of y_{it} conditional on individual intercept now denoted by d_i, $\boldsymbol{X}_i$, and $\boldsymbol{y}_{i,1:t-1}$ as

$$p(y_{it}|d_i, \boldsymbol{X}_i, \boldsymbol{y}_{i,1:t-1}) = p(y_{it}|d_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1}), \tag{13}$$

retaining the assumption that previous covariates and responses before $y_{i,t-1}$ do not affect y_{it}. Note that, differently from (10), the conditioning set on the rhs includes the first-order leads of $\boldsymbol{x}_{it}$. Moreover, we use a different symbol for the unobserved individual intercept that, as will be clear in the following, is related to the individual parameter d_i. The formulation can easily be extended to include an arbitrary number of leads $\boldsymbol{X}_{i,t:t+H}$, with $H \leq T-3$, so that we retain at least two observations, which is necessary for inference (see Section 4). However, we do not explicitly consider this extension because, while being rather obvious, it strongly complicates the following exposition.[2] Following the discussion in Chamberlain (1984) and the suggestion in Wooldridge (2010, 15.8.2) on testing the strict exogeneity assumption, a test for noncausality can be derived by specifying the model as

$$p(y_{it}|d_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1}) = g^{-1}(d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma), \quad t = 2, \ldots, T-1,$$

where $g^{-1}(\cdot)$ is an inverse link function. It is worth noting that the null hypothesis $H_0 : \boldsymbol{\nu} = 0$ corresponds to condition S', and then to Granger noncausality G. The identification of $\boldsymbol{\beta}$ and γ in presence of departures from noncausality requires further assumptions that lead to the formulation here proposed. In particular, we rely on the logit formulation

$$p(y_{it}|d_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1}) = \frac{\exp\left[y_{it}\left(d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{it-1}\gamma\right)\right]}{1 + \exp\left(d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma\right)}. \tag{14}$$

Under a particular, very relevant, case this formulation is justified according to the following arguments.

First of all, denote the conditional density of the distribution of the covariate vector $\boldsymbol{x}_{i,t+1}$ as

$$f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{X}_{i,1:t}, \boldsymbol{y}_{i,1:t}) = f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it}), \quad t = 1, \ldots, T-1, \tag{15}$$

where $\boldsymbol{\xi}_i$ is a column vector of time-fixed effects and the presence of y_{it} allows for feedback effects.[3] Then the logit for the distribution y_{it} conditional on c_i, $\boldsymbol{\xi}_i$, $\boldsymbol{X}_{i,t:t+1}$, and $y_{i,t-1}$ is

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

$$\log\frac{p(y_{it}=1|c_i, \boldsymbol{x}_{it}, y_{i,t-1})f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it}=1)}{p(y_{it}=0|c_i, \boldsymbol{x}_{it}, y_{i,t-1})f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it}=0)}, \tag{16}$$

where the presence of time-fixed effects in the conditioning sets for y_{it} and $\boldsymbol{x}_{it}$ is determined by (13) and (15).[4] Furthermore, we assume that the probability of y_{it} conditional on c_i, $\boldsymbol{x}_{it}$, $y_{i,t-1}$ has the dynamic logit formulation expressed in (11) so that the above expression

[2] Chamberlain (1984) reports an empirical example where the linear index function of a logit model corresponds to the lhs of S in (3), 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.

[3] In assumption (15) we maintain the same first-order dynamic as for (13). Nevertheless the assumptions on the conditioning set on the right-hand-side can be relaxed to include more lags of $\boldsymbol{x}_{it}$ and y_{it}.

[4] Notice that the extension of (13) to a number of leads $1 < H \leq T-3$ requires to rewrite the conditional density of covariates as $\prod_{h=1}^{H} r(\boldsymbol{x}_{i,t+h}|\boldsymbol{\xi}_i, \boldsymbol{x}_{i,t+h-1}, y_{it} = z)$, with $z = 0, 1$.

becomes

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

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 following exponential family:

$$f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i,\boldsymbol{x}_{it},y_{it}=z)=\frac{\exp[\boldsymbol{x}_{i,t+1}'(\boldsymbol{\xi}_i+\boldsymbol{\eta}_z)]h(\boldsymbol{x}_{i,t+1};\boldsymbol{\sigma})}{K(\boldsymbol{\xi}_i+\boldsymbol{\eta}_z;\boldsymbol{\sigma})},\quad t=1,\dots,T-1,\ z=0,1, \tag{17}$$

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 holds, the proof of which is trivial.

Theorem 3. *Under assumptions* (11) *and* (17)*, we have*

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

$$d_i+\boldsymbol{x}_{it}'\boldsymbol{\beta}+\boldsymbol{x}_{i,t+1}'\boldsymbol{\nu}+y_{i,t-1}\gamma,$$

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* (14) *holds.*

Two cases satisfying (17) are for continuous covariates having multivariate normal distribution with common variance-covariance matrix and the case of binary covariates. More precisely, in the first case 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 (17) 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 $\boldsymbol{\Sigma}$ go in $\boldsymbol{\psi}$. Regarding 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 Bernoulli distribution with success probability

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

where k is the number of covariates. In the other cases, when (17) does not hold, 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 (16) which is the most natural solution to maintain an acceptable level of simplicity.

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 (14) and that directly compares with (12). For all i, the distribution at issue is

$$p(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = \qquad (18)$$
$$\frac{\exp\left[y^*_{i+}d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{it-1}\gamma\right)\right]}{\prod_{t=2}^{T-1}\left[1 + \exp\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\gamma\right)\right]}.$$

where $y^*_{i+} = \sum_{t=2}^{T-1} y_{it}$. In particular, model (18) reduces to the dynamic logit (12) under the null hypothesis of noncausality $H_0 : \boldsymbol{\nu} = \boldsymbol{0}$, if the probability in (12) is conditioned on y_{iT} and with different individual intercepts.

The parameters in (18) can be estimated by either a random- or fixed-effects approach, keeping in mind that a (correlated) random-effects strategy (Mundlak, 1978; Chamberlain, 1984) requires the predetermined covariates in $\boldsymbol{x}_{it}$ to be independent of d_i. As this assumption may often be hardly tenable, in the next section we discuss a fixed-effects estimation approach, first put forward by Bartolucci and Nigro (2012) and here adapted to the present case.

4 Fixed-effects estimation

With fixed-T panel data, a fixed-effects approach to the estimation of the parameters of the standard logit model is based on the maximization of the conditional likelihood given suitable sufficient statistics for the incidental parameters. The conditional estimator is common practice for static binary panel data models (Chamberlain, 1980), whereas, for the dynamic logit model, a sufficient statistic can only be derived in special cases: in absence of covariates with $T = 3$ (Chamberlain, 1985); with covariates on the basis of a weighted conditional log-likelihood, although the estimator is consistent only under certain 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 approximate the dynamic logit with a QE model (Cox, 1972; Bartolucci and Nigro, 2010), which admits a sufficient statistic 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 propose to adopt a PCML estimator for the model parameters. In the following, we extend the approximating QE model to accommodate the parametrization of the proposed model formulation in (18).

4.1 Approximating model

The approximating model for (18) is derived by taking a linearization of the log-probability of the latter, similar to that used in Bartolucci and Nigro (2012), that is,

$$\log p(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = y_{i+}^* d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma\right) -$$

$$\sum_{t=2}^{T-1} \log\left[1 + \exp\left(d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma\right)\right]. \tag{19}$$

The term that is nonlinear in the parameters is approximated by a first-order Taylor series expansion around $d_i = \bar{d}_i$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, $\boldsymbol{\nu} = \bar{\boldsymbol{\nu}}$, and $\gamma = 0$, leading to

$$\sum_{t=2}^{T-1} \log\left[1 + \exp\left(d_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma\right)\right] \approx$$

$$\sum_{t=2}^{T-1} \left[1 + \exp\left(\bar{d}_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}} + \boldsymbol{x}_{i,t+1}'\bar{\boldsymbol{\nu}}\right)\right] +$$

$$\sum_{t=2}^{T-1} q_{it}\left[d_i - \bar{d}_i + \boldsymbol{x}_{it}'\left(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}\right) + \boldsymbol{x}_{i,t+1}'\left(\boldsymbol{\nu} - \bar{\boldsymbol{\nu}}\right)\right] + \sum_{t=2}^{T-1} q_{it} y_{i,t-1}\gamma, \tag{20}$$

where

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

Since only the last sum in (20) depends on $\boldsymbol{y}_{i,2:T-1}$, we can substitute (20) in (19) and obtain the approximation of the joint probability (18) that gives the following QE model

$$p^*(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = \frac{\exp\left[y_{i+}^* d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu}\right) + \sum_t (y_{it} - q_{it}) y_{i,t-1}\gamma\right]}{\sum_{\boldsymbol{z}_{2:T-1}} \exp\left[z_+^* d_i + \sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu}\right) + \sum_t (z_t - q_{it}) z_t\gamma\right]}, \tag{21}$$

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

The joint probability in (21) is closely related to the probability of the response configuration $\boldsymbol{y}_{i,2:T-1}$ in the true model in (18). In particular, the approximating QE and the proposed true model share the properties summarized by the following theorem that can be proved along the lines of Bartolucci and Nigro (2010):[5]

[5] Results (*ii*) and (*iii*) can easily be derived by extending to the present case Theorem 1 in Bartolucci

Theorem 4. *For $i = 1, \dots, n$:*

(i) *In the case of $\gamma = 0$, the joint probability $p^*(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT})$ does not depend on $y_{i,t-1}$ or on q_{it}, and both the true* (18) *and approximating model* (21)*, correspond to the following static logit model*

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

(ii) *y_{it} is conditionally independent of $\boldsymbol{y}_{i,1:t-2}$ given d_i, $\boldsymbol{X}_i$, and $y_{i,t-1}$, for $t = 2, \dots, T$.*

(iii) *Under both models, the parameter γ has the same interpretation in terms of log-odds ratio between the responses y_{it} and $y_{i,t-1}$, for $t = 2, \dots, T-1$:*

$$\log \frac{p^*(y_{it} = 1|d_i, \boldsymbol{X}_i, y_{i,t-1} = 1)}{p^*(y_{it} = 0|d_i, \boldsymbol{X}_i, y_{i,t-1} = 1)} - \log \frac{p^*(y_{it} = 1|d_i, \boldsymbol{X}_i, y_{i,t-1} = 0)}{p^*(y_{it} = 0|d_i, \boldsymbol{X}_i, y_{i,t-1} = 0)} = \gamma.$$

The nice feature of the QE model in (21) is that it admits sufficient statistics for the incidental parameters d_i, which are the total scores y^*_{i+} for $i = 1, \dots, n$. The probability of $\boldsymbol{y}_{i,2:T-1}$, conditional on $\boldsymbol{X}_i$, y_{i1}, y_{iT}, and y^*_{i+}, for the approximating model is then

$$p^*\left(\boldsymbol{y}_{i,2:T-1}|\boldsymbol{X}_i, y_{i1}, y_{iT}, y^*_{i+}\right) = \frac{\exp\left[\sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(y_{it} - q_{it})y_{i,t-1}\gamma\right]}{\sum_{\substack{\boldsymbol{z}_{2:T-1} \\ z^*_+ = y^*_{i+}}} \exp\left[\sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(z_t - q_{it})z_{t-1}\gamma\right]}, \tag{22}$$

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

4.2 Pseudo conditional maximum likelihood estimator

The formulation of the conditional log-likelihood for (22) relies on the fixed quantities q_{it}, that are based on a preliminary estimation of the parameters associated with the covariate and of the individual effects. Let $\boldsymbol{\phi} = (\boldsymbol{\beta}', \boldsymbol{\nu}')'$ be the vector collecting all the regression parameters and $\boldsymbol{\theta} = (\boldsymbol{\phi}', \boldsymbol{\gamma}')'$. The estimation approach is based on two-steps:

and Nigro (2012), that clarifies the connection between the QE and the dynamic logit model.

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

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

which can be maximized by a Newton-Raphson algorithm.

2. The parameter vector $\boldsymbol{\theta}$ is estimated by maximizing the conditional log-likelihood of (22), that can be written as

$$\begin{aligned}\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \sum_i 1\{0 < y_{it} < (T-2)\}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}), \qquad (23)\\ \ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \log p^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}(\boldsymbol{y}_{i,2:T-1}|\boldsymbol{X}_i, y_{i1}, y_{i1}, y_{i+}^*).\end{aligned}$$

The resulting $\hat{\boldsymbol{\theta}}$ is the pseudo conditional maximum likelihood estimator.

Function $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ may be maximized by Newton-Raphson using the score and observed information matrix reported below (Section 4.2.1). We also illustrate how to derive standard errors for the two-step estimator (Section 4.2.2). We leave out of the exposition the asymptotic properties of the PCML estimator, which can be derived along the same lines as in Bartolucci and Nigro (2012).

4.2.1 Score and information matrix

In order to write the score and information matrix for $\boldsymbol{\theta}$, it is convenient to rewrite $\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ as

$$\begin{aligned}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) = {} & \boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1})'\boldsymbol{A}^*(\boldsymbol{X}_i)'\boldsymbol{\theta} - \\ & \log \sum_{\substack{\boldsymbol{z}_{2:T-1}\\ z_+^* = y_{i+}^*}} \exp\left[\boldsymbol{u}^*(\boldsymbol{z}_{i,1:T-1})'\boldsymbol{A}^*(\boldsymbol{X}_i)'\boldsymbol{\theta}\right], \qquad (24)\end{aligned}$$

where the notation $\boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1})$ is used to stress that $\boldsymbol{u}^*$ is a function of both the initial value y_{i1} and the response configuration $\boldsymbol{y}_{i,2:T-1}$; similarly $\boldsymbol{u}^*(\boldsymbol{z}_{i,1:T-1})$ is a function of y_{i1}

and $\boldsymbol{z}_{2:T-1}$, since $z_1 = y_{i1}$ as in (21). Moreover $\boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1})$ and $\boldsymbol{A}^*(\boldsymbol{X}_i)$ in (24) are

$$\begin{aligned} \boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1}) &= \left(\boldsymbol{y}'_{i,2:T-1}, \sum_{t=2}^{T-1}(y_{it} - q_{it})y_{i,t-1}\right)' \\ \boldsymbol{A}^*(\boldsymbol{X}_i) &= \left(\begin{array}{cc} \boldsymbol{X}_{i,2:T} & 0 \\ \boldsymbol{0}' & 1 \end{array}\right), \end{aligned} \tag{25}$$

where $\boldsymbol{X}_{i,2:T}$ is a matrix of $T-1$ rows and $2k$ columns, with k the number of covariates and typical row $\boldsymbol{x}'_{i,t:t+1}$, while $\mathbf{0}$ is column vector of zeros having a suitable dimension.[6] Using the above notation, the score $\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) = \nabla_{\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ and the observed information matrix $\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) = -\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ are

$$\begin{aligned} \boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{A}^*(\boldsymbol{X}_i)\{\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1}) - \\ &\qquad \mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, , y_{iT}, y_{i+}^*\right]\}, \end{aligned} \tag{26}$$

and

$$\begin{aligned} \boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{A}^*(\boldsymbol{X}_i) \times \\ &\qquad \mathrm{V}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right]\boldsymbol{A}^*(\boldsymbol{X}_i)', \end{aligned} \tag{27}$$

where the conditional expected value and variance are defined as

$$\begin{aligned} &\mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right] = \\ &\qquad \sum_{\substack{\boldsymbol{z}_{H+1:T-H} \\ z_+^* = y_{i+}^*}} \boldsymbol{u}^*(\boldsymbol{z}_{i,2:T-2})p^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left(\boldsymbol{z}_{i,2:T-2}|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right), \end{aligned}$$

and

$$\begin{aligned} &\mathrm{V}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, \boldsymbol{y}_{i,1:H}, y_{i+}^*\right] = \\ &\qquad \mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})'|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right] - \\ &\qquad \mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right]\mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right]'. \end{aligned}$$

Following the results in Bartolucci and Nigro (2012), which can be applied directly to

[6] In order to clarify the structure of $\boldsymbol{A}^*(\boldsymbol{X}_i)$, consider the simple case of $T = 4$ time occasions and one covariate. Then

$$\boldsymbol{A}^*(\boldsymbol{X}_i) = \left(\begin{array}{ccc} x_{i2} & x_{i3} & 0 \\ x_{i3} & x_{i4} & 0 \\ 0 & 0 & 1 \end{array}\right).$$

the present case, $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is always concave and $\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is almost surely positive definite.[7] Then $\hat{\boldsymbol{\theta}}$ that maximizes $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is found at convergence of the standard Newton-Raphson algorithm.

4.2.2 Standard errors

The computation of standard errors must take into account the first step estimation of $\bar{\boldsymbol{\phi}}$. As Bartolucci and Nigro (2012) we also rely on the GMM approach (Hansen, 1982) and cast the estimating equations as

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

where $\boldsymbol{m}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta})$ contains the score vectors of the first step, $\nabla_{\bar{\boldsymbol{\phi}}}\ell_i(\bar{\boldsymbol{\phi}})$, and of the second step, $\nabla_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$. Then the GMM estimator is $(\tilde{\boldsymbol{\phi}}', \hat{\boldsymbol{\theta}}')'$ and its variance-covariance matrix can be estimated as

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

where

$$\begin{aligned}
\boldsymbol{S}(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{m}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta})\boldsymbol{m}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta})', \\
\boldsymbol{H}(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{H}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}).
\end{aligned}$$

Matrix $\boldsymbol{H}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta})$ is composed of four blocks as follows:

$$\boldsymbol{H}_i(\bar{\boldsymbol{\phi}}, \boldsymbol{\theta}) = \begin{pmatrix} \nabla_{\bar{\boldsymbol{\phi}}\bar{\boldsymbol{\phi}}}\ell_i(\bar{\boldsymbol{\phi}}) & \boldsymbol{0} \\ \nabla_{\boldsymbol{\theta}\bar{\boldsymbol{\phi}}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) & \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) \end{pmatrix}.$$

The north-west block is expressed as

$$\nabla_{\bar{\boldsymbol{\phi}}\bar{\boldsymbol{\phi}}}\ell_i(\bar{\boldsymbol{\phi}}) = \boldsymbol{X}_{i,2:T}\mathrm{V}_{\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{iT}, y_{i+}^*\right]\boldsymbol{X}_{i,2:T}',$$

where $\boldsymbol{X}_{i,2:T}$ is defined in (25) and $\mathrm{V}_{\bar{\boldsymbol{\phi}}}$ is the conditional variance in the static logit model. Moreover, $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is equal to $-\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$; see definition (27). Finally, the derivation of $\nabla_{\boldsymbol{\theta}\bar{\boldsymbol{\phi}}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is not straightforward and we therefore rely on the numerical derivative of (26) with respect to $\bar{\boldsymbol{\phi}}$.

[7] See Bartolucci and Nigro (2012), Section 5, Theorem 2.

5 Simulation study

In this section we describe the design and illustrate the main results of the simulation study we used to investigate about the final sample properties of the PCML estimator for the parameters of the proposed model formulation. In the first part of the study, the main focus is on the performance under substantial departures from noncausality, which we obtain by a non-zero effect from the past values of the binary dependent variable on the present value of the covariate. In the second part, we compare the PCML estimator of (18) with an alternative ML random-effects estimator for the same model, based on the proposal by Wooldridge (2005) to account for the initial condition problem.

5.1 Simulation design

The simulation study is based on samples drawn from a dynamic logit model, where the linear index specification includes the lagged dependent variable, one explanatory variable x_{it} possibly predetermined, one strictly exogenous variable v_{it}, and individual unobserved heterogeneity. The model assumes that

$$y_{it} = 1\{c_i + \beta x_{it} - 0.5 v_{it} + \gamma y_{it-1} + \varepsilon_{it} \geq 0\}, \tag{28}$$

for $i = 1, \ldots, n$, $t = 2, \ldots, T$, with initial condition

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

In the considered scenarios, the error terms ε_{it}, $t = 1, \ldots, T$, follow a logistic distribution with zero mean and variance equal to $\pi^2/3$ and the individual specific intercepts c_i are allowed to be correlated with x_{it} and v_{it}.

We consider a benchmark design and some extensions that are characterized by different choices for the distribution of the explanatory variable x_{it}. The general formulation is

$$\begin{aligned} x_{it} &= w(\xi_i + x_{it}^* + \psi v_{it} + \eta y_{it-1}), \\ x_{it}^* &\sim N(0, \pi^2/3), \end{aligned} \tag{29}$$

for $t = 2, \ldots, T$, the initial value is $x_{i1} = w(\xi_i + x_{i1}^* + \psi v_{i1})$ with x_{i1}^* being again a zero mean normal with variance $\pi^2/3$, and $v_{it} = \xi_i + v_{it}^*$, for $t = 1, \ldots, T$, where v_{it}^* is also $N(0, \pi^2/3)$. The parameter η governs the violation of S', stated in Section 2, and it takes value $\eta = 0$ under the assumption of noncausality, with $\eta \neq 0$ otherwise. In our benchmark design, we let $w(\cdot)$ be the identity function and $\psi = 0$, so that assumption (17) is satisfied and the model of Theorem 3 holds. We also consider two alternative designs where (17) does not hold and the model formulated in Theorem 3 is an approximation: first, we let $w(\cdot)$ be an indicator function so that x_{it} becomes a binary covariate with a normally distributed error term, with $p(x_{it}=1|\xi_i, v_{it}, y_{i,t-1}) = \Phi(\xi_i + x^*_{it} + \psi v_{it} + \eta y_{i,t-1})$, where $\Phi(\cdot)$ is the standard normal cdf and therefore does not belong to the exponential family in (17); secondly, we let the $w(\cdot)$ be the identity function and set $\psi = 0.5$ in order for x_{it} to depend on other time varying covariates.

Based on x^*_{it}, the individual intercepts c_i and ξ_i are derived as

$$\begin{aligned} c_i &= (1/T)\sum_{t=1}^{4} x^*_{it}, \\ \xi_i &= \varpi\, c_i + \sqrt{1-\varpi^2}u_{it}, \end{aligned} \tag{30}$$

with $\varpi = 0.5$, $u_{it} \sim N(0,1)$ and for $i = 1,\dots,n$. This way, the generating model admits a correlation between the covariates and the individual-specific intercepts and dependence between the unobserved heterogeneity in both processes for y and x.

In most economic applications, the parameters of interest are γ, measuring the state dependence, and the regression coefficient β. Based on the generating model (28), we ran experiments for scenarios with $\gamma = 0, 1$ and $\beta = 0, -1$. We examine violations of noncausality by setting $\eta = -1$, compared with the same scenarios with $\eta = 0$. The chosen values for β, γ, and η are consistent with likely situations in practice that relate, for instance, to the feedback effect of past employment on present child birth when analyzing female labor supply (see also Mosconi and Seri, 2006, for a related application). Notice that here we are implicitly assuming 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. The sample sizes considered are $n = 500, 1000$ for $T = 4, 8$. The number of Monte Carlo replications is 1000.

5.2 Main results

Tables 1–6 report the main results of our simulation study. Tables 1–4 show the results for the benchmark design, under which the covariate x_{it}, generated as in (29), is normally distributed, with $w(\cdot)$ being the identity function, and $\psi = 0$, for all the combinations of the chosen values of β and γ. Tables 5 and 6 report the simulation results for the two extensions of our benchmark design, under which x_{it} is generated as a binary variable and with a dependence on the time varying covariate v_{it}, respectively, for $\beta = -1$, $\gamma = 1$, and $\eta = 0, -1$.

For each scenario, we investigate the finite sample performance of the PCML estimator in Section 4 for the proposed formulation (18) in two cases representing the null and alternative hypotheses of noncausality described by S' in Section 2: PCML$_1$ denotes the PCML estimator for the parameters in (18); PCML$_0$ denotes the estimator of (18) with the constraint $\nu = 0$. 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 $H_0 : \hat{\beta} = \beta$, and $H_0 : \hat{\gamma} = \gamma$. Finally we report the t-tests at the 5% nominal size for noncausality, $H_0 : \nu = 0$. We expect PCML$_0$ to yield biased estimators when $\eta \neq 0$ since, following S', the lead of x_{it} is omitted from the model specification. We limit the discussion to the estimation of β and γ, which are likely to be the parameters of main interest in applications. Results concerning the other model parameters are available upon request.

Table 1 summarizes the simulation results for our benchmark design and $\beta = \gamma = 0$. With $\eta = 0$, that is, in absence of feedback effects, the mean bias and median bias are always negligible, whereas the MAE and RMSE decrease with both n and T for the two models considered. While the same considerations hold for PCML$_1$ when $\eta = -1$, the PCML estimators of β provided by PCML$_0$ is severely biased and leads to misleading inference, although this pattern is alleviated for $T = 8$. The same patterns are shown in Table 2, where β is equal to -1. Moreover, the t-test for $H_0 : \nu = 0$ always attains its nominal size and exhibits strong empirical power in all the scenarios with $\eta = -1$

Tables 3 and 4 summarize simulation results for the same designs when $\gamma = 1$. They depict similar situations to those in Tables 1 and 2, with the exception of the bias of γ, that slightly increases. In fact, the performance of the PCML estimator may be especially sensitive to the degree of state dependence in the generated samples. A high value of γ leads to a reduction of the actual sample size via the indicator function in (23) and represents a large deviation from the approximating point by which (20) is derived. Nevertheless, Bartolucci and Nigro (2012) show that the bias and root-mean square error of PCML estimator of γ in the dynamic logit model decrease at a rate close to $\sqrt{n}$ and as T grows also for γ moving away from 0.

Tables 5 and 6 report the simulation results for two departures from the benchmark design: Table 5 refers to a binary covariate generated by a normal link function, while Table 6 refers to a normally distributed covariate depending on the time-varying covariate v_{it} (see Section 5.1 for details). These exercises are meant to investigate the properties of the PCML estimator when assumption (17) does not hold and the model formulated in Theorem 3 just embeds a linear approximation of (17). When the covariate is binary, the bias of the PCML$_1$ estimator of β and γ is always negligible. As for efficiency, the RMSE and MAE are slightly higher for β, although they decrease with both n and T (see Table 5). On the other hand, the results for $\psi = 0.5$ in Table 6 mirror closely those in Table 4, except for a larger bias with $T = 4$.

Table 1: Normally distributed covariate, $\beta = 0$, $\gamma = 0$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML$_1$	-0.003	0.072	-0.003	0.048	0.052	-0.026	0.305	-0.031	0.210	0.039	0.051
PCML$_0$	-0.001	0.060	0.001	0.039	0.045	-0.027	0.302	-0.025	0.209	0.036	
					$n = 500$, $T = 8$						
PCML$_1$	-0.000	0.027	0.000	0.018	0.066	0.003	0.106	0.002	0.073	0.055	0.037
PCML$_0$	-0.000	0.027	-0.000	0.018	0.062	0.003	0.106	0.002	0.073	0.056	
					$n = 1000$, $T = 4$						
PCML$_1$	0.000	0.051	-0.000	0.034	0.051	0.002	0.224	0.009	0.143	0.055	0.050
PCML$_0$	-0.000	0.043	-0.001	0.029	0.052	0.002	0.223	0.010	0.143	0.052	
					$n = 1000$, $T = 8$						
PCML$_1$	0.001	0.019	0.001	0.012	0.048	0.000	0.074	-0.002	0.048	0.053	0.055
PCML$_0$	0.001	0.018	0.001	0.012	0.053	0.000	0.074	-0.002	0.048	0.053	
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML$_1$	0.002	0.078	-0.001	0.054	0.042	-0.013	0.338	-0.009	0.224	0.045	0.984
PCML$_0$	0.155	0.167	0.154	0.154	0.694	0.138	0.346	0.152	0.236	0.057	
					$n = 500$, $T = 8$						
PCML$_1$	-0.003	0.027	-0.002	0.018	0.047	-0.000	0.112	-0.000	0.076	0.044	1.000
PCML$_0$	0.048	0.054	0.048	0.048	0.498	0.053	0.115	0.049	0.078	0.078	
					$n = 1000$, $T = 4$						
PCML$_1$	-0.002	0.053	-0.002	0.037	0.051	-0.003	0.245	-0.002	0.166	0.055	1.000
PCML$_0$	0.149	0.155	0.149	0.149	0.935	0.149	0.275	0.153	0.195	0.089	
					$n = 1000$, $T = 8$						
PCML$_1$	-0.003	0.020	-0.004	0.014	0.071	0.004	0.080	0.003	0.055	0.046	1.000
PCML$_0$	0.048	0.052	0.048	0.048	0.795	0.057	0.092	0.056	0.063	0.129	

Table 2: Normally distributed covariate, $\beta = -1$, $\gamma = 0$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML_1	-0.049	0.178	-0.027	0.106	0.039	0.037	0.482	0.028	0.325	0.056	0.055
PCML_0	-0.039	0.165	-0.020	0.102	0.048	0.033	0.473	0.018	0.318	0.056	
					$n = 500$, $T = 8$						
PCML_1	-0.007	0.049	-0.005	0.034	0.057	-0.006	0.135	-0.000	0.095	0.049	0.045
PCML_0	-0.007	0.049	-0.004	0.033	0.056	-0.006	0.134	-0.001	0.094	0.053	
					$n = 1000$, $T = 4$						
PCML_1	-0.019	0.117	-0.005	0.075	0.043	0.005	0.309	0.010	0.219	0.041	0.037
PCML_0	-0.015	0.111	-0.007	0.073	0.046	0.006	0.307	0.007	0.222	0.042	
					$n = 1000$, $T = 8$						
PCML_1	-0.001	0.035	0.001	0.023	0.051	0.005	0.090	0.006	0.060	0.040	0.056
PCML_0	-0.001	0.035	0.001	0.022	0.055	0.005	0.090	0.007	0.060	0.041	
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML_1	-0.058	0.208	-0.037	0.122	0.058	-0.015	0.501	-0.020	0.333	0.051	0.808
PCML_0	0.122	0.199	0.138	0.158	0.222	0.045	0.474	0.058	0.317	0.050	
					$n = 500$, $T = 8$						
PCML_1	-0.006	0.055	-0.005	0.035	0.049	0.002	0.148	0.002	0.101	0.058	1.000
PCML_0	0.047	0.069	0.048	0.052	0.194	-0.097	0.170	-0.095	0.122	0.112	
					$n = 1000$, $T = 4$						
PCML_1	-0.027	0.134	-0.018	0.082	0.060	-0.003	0.340	-0.003	0.224	0.049	0.981
PCML_0	0.140	0.177	0.148	0.150	0.330	0.055	0.325	0.043	0.213	0.051	
					$n = 1000$, $T = 8$						
PCML_1	-0.003	0.039	-0.003	0.027	0.056	0.007	0.101	0.007	0.069	0.055	1.000
PCML_0	0.050	0.061	0.051	0.051	0.311	-0.091	0.133	-0.091	0.096	0.172	

Table 3: Normally distributed covariate, $\beta = 0$, $\gamma = 1$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.002	0.079	0.001	0.051	0.033	-0.003	0.418	-0.000	0.289	0.063	0.040
PCML$_0$	-0.000	0.069	-0.003	0.046	0.025	-0.010	0.412	-0.013	0.288	0.060	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.002	0.027	-0.003	0.018	0.049	0.005	0.113	0.004	0.076	0.048	0.052
PCML$_0$	-0.002	0.027	-0.003	0.017	0.049	0.005	0.113	0.003	0.075	0.046	
					$n = 1000$,	$T = 4$					
PCML$_1$	-0.003	0.054	-0.003	0.037	0.031	-0.025	0.279	-0.029	0.195	0.052	0.035
PCML$_0$	-0.002	0.048	-0.000	0.032	0.045	-0.029	0.277	-0.033	0.193	0.049	
					$n = 1000$,	$T = 8$					
PCML$_1$	-0.001	0.020	-0.000	0.014	0.051	-0.001	0.080	-0.006	0.054	0.049	0.059
PCML$_0$	-0.001	0.020	-0.000	0.014	0.051	-0.002	0.080	-0.005	0.054	0.048	
$\eta = -1$											
					$n = 500$,	$T = 4$					
PCML$_1$	0.006	0.085	0.008	0.056	0.037	-0.004	0.441	-0.016	0.297	0.050	0.894
PCML$_0$	0.143	0.157	0.143	0.143	0.520	0.147	0.442	0.140	0.281	0.055	
					$n = 500$,	$T = 8$					
PCML$_1$	0.007	0.031	0.006	0.021	0.065	0.006	0.125	0.003	0.084	0.057	1.000
PCML$_0$	0.018	0.032	0.017	0.022	0.104	0.008	0.114	0.002	0.078	0.055	
					$n = 1000$,	$T = 4$					
PCML$_1$	0.004	0.060	0.005	0.042	0.039	-0.001	0.301	-0.002	0.191	0.059	0.992
PCML$_0$	0.139	0.147	0.137	0.137	0.815	0.148	0.323	0.147	0.225	0.075	
					$n = 1000$,	$T = 8$					
PCML$_1$	0.004	0.020	0.003	0.013	0.059	0.002	0.089	0.004	0.060	0.055	1.000
PCML$_0$	0.015	0.023	0.014	0.016	0.118	0.005	0.082	0.003	0.056	0.058	

Table 4: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.030	0.207	0.003	0.120	0.056	0.056	0.571	0.032	0.365	0.038	0.056
PCML$_0$	-0.027	0.190	-0.001	0.106	0.052	0.045	0.560	0.035	0.360	0.038	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.007	0.052	-0.005	0.036	0.048	0.005	0.143	0.006	0.092	0.059	0.056
PCML$_0$	-0.006	0.052	-0.003	0.036	0.048	0.005	0.142	0.004	0.093	0.055	
					$n = 1000$,	$T = 4$					
PCML$_1$	0.006	0.124	0.012	0.085	0.063	0.009	0.393	0.001	0.267	0.050	0.043
PCML$_0$	0.000	0.116	0.007	0.077	0.050	0.012	0.389	0.001	0.265	0.044	
					$n = 1000$,	$T = 8$					
PCML$_1$	-0.001	0.036	-0.001	0.024	0.047	0.009	0.100	0.011	0.064	0.057	0.058
PCML$_0$	-0.000	0.036	-0.000	0.024	0.047	0.009	0.099	0.009	0.065	0.056	
$\eta = -1$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.031	0.211	-0.002	0.133	0.045	0.035	0.632	0.032	0.392	0.041	0.509
PCML$_0$	0.123	0.219	0.148	0.175	0.185	0.053	0.590	0.055	0.386	0.045	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.003	0.059	0.001	0.041	0.052	-0.020	0.158	-0.021	0.108	0.052	1.000
PCML$_0$	0.022	0.060	0.025	0.042	0.084	-0.150	0.211	-0.147	0.155	0.186	
					$n = 1000$,	$T = 4$					
PCML$_1$	0.012	0.139	0.025	0.095	0.057	0.018	0.405	0.012	0.269	0.035	0.809
PCML$_0$	0.151	0.193	0.165	0.168	0.334	0.045	0.391	0.042	0.261	0.037	
					$n = 1000$,	$T = 8$					
PCML$_1$	0.003	0.043	0.005	0.029	0.059	-0.016	0.113	-0.015	0.079	0.046	1.000
PCML$_0$	0.027	0.048	0.029	0.035	0.130	-0.145	0.180	-0.144	0.144	0.299	

Table 5: Binary covariate, $\beta = -1$, $\gamma = 1$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.007	0.352	-0.005	0.242	0.040	0.005	0.398	0.009	0.263	0.049	0.045
PCML$_0$	-0.011	0.309	0.001	0.210	0.038	-0.003	0.390	0.004	0.260	0.048	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.010	0.116	-0.010	0.078	0.050	0.000	0.113	-0.002	0.076	0.053	0.060
PCML$_0$	-0.008	0.115	-0.009	0.076	0.049	0.000	0.113	-0.001	0.076	0.051	
					$n = 1000$,	$T = 4$					
PCML$_1$	0.019	0.238	0.023	0.160	0.042	-0.018	0.279	-0.029	0.187	0.060	0.045
PCML$_0$	-0.000	0.211	0.003	0.140	0.040	-0.019	0.277	-0.033	0.187	0.057	
					$n = 1000$,	$T = 8$					
PCML$_1$	-0.009	0.080	-0.012	0.054	0.049	0.004	0.079	0.002	0.054	0.040	0.065
PCML$_0$	-0.008	0.079	-0.010	0.053	0.047	0.004	0.079	0.001	0.054	0.040	
$\eta = -1$											
					$n = 500$,	$T = 4$					
PCML$_1$	0.022	0.364	0.038	0.236	0.044	0.001	0.409	-0.007	0.278	0.048	0.579
PCML$_0$	0.432	0.528	0.447	0.449	0.309	0.042	0.399	0.029	0.267	0.052	
					$n = 500$,	$T = 8$					
PCML$_1$	0.008	0.121	0.005	0.083	0.047	-0.003	0.116	-0.009	0.080	0.048	1.000
PCML$_0$	0.049	0.124	0.046	0.080	0.074	-0.024	0.114	-0.027	0.083	0.049	
					$n = 1000$,	$T = 4$					
PCML$_1$	0.044	0.265	0.063	0.185	0.048	-0.022	0.290	-0.032	0.193	0.052	0.883
PCML$_0$	0.447	0.494	0.450	0.451	0.553	0.029	0.283	0.018	0.189	0.055	
					$n = 1000$,	$T = 8$					
PCML$_1$	0.013	0.088	0.014	0.057	0.063	-0.001	0.081	0.002	0.055	0.043	1.000
PCML$_0$	0.053	0.098	0.052	0.067	0.108	-0.022	0.081	-0.019	0.054	0.057	

Table 6: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $\psi = 0.5$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML$_1$	-0.075	0.278	-0.037	0.140	0.043	0.103	0.774	0.111	0.469	0.049	0.058
PCML$_0$	-0.044	0.222	-0.015	0.125	0.039	0.077	0.708	0.073	0.447	0.038	
					$n = 500$, $T = 8$						
PCML$_1$	-0.006	0.058	-0.004	0.036	0.054	0.007	0.154	0.008	0.101	0.032	0.056
PCML$_0$	-0.004	0.057	-0.001	0.036	0.053	0.005	0.152	0.006	0.098	0.035	
					$n = 1000$, $T = 4$						
PCML$_1$	-0.017	0.158	-0.009	0.099	0.064	0.013	0.491	-0.008	0.321	0.038	0.046
PCML$_0$	-0.008	0.144	0.004	0.091	0.063	0.009	0.475	-0.024	0.316	0.034	
					$n = 1000$, $T = 8$						
PCML$_1$	-0.002	0.042	0.001	0.027	0.049	0.015	0.113	0.013	0.073	0.049	0.049
PCML$_0$	-0.001	0.041	0.001	0.027	0.047	0.015	0.112	0.013	0.074	0.051	
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML$_1$	-0.115	0.372	-0.045	0.170	0.062	0.087	0.970	0.022	0.527	0.071	0.408
PCML$_0$	0.092	0.257	0.132	0.184	0.164	0.059	0.810	0.008	0.475	0.065	
					$n = 500$, $T = 8$						
PCML$_1$	-0.002	0.066	-0.001	0.044	0.057	-0.001	0.183	-0.000	0.119	0.061	1.000
PCML$_0$	0.027	0.067	0.028	0.048	0.092	-0.107	0.200	-0.101	0.133	0.115	
					$n = 1000$, $T = 4$						
PCML$_1$	-0.027	0.191	-0.001	0.119	0.055	0.032	0.538	0.029	0.345	0.050	0.690
PCML$_0$	0.133	0.203	0.151	0.167	0.248	0.054	0.503	0.053	0.318	0.048	
					$n = 1000$, $T = 8$						
PCML$_1$	0.001	0.046	0.002	0.032	0.060	-0.014	0.126	-0.014	0.084	0.055	1.000
PCML$_0$	0.029	0.053	0.030	0.037	0.121	-0.118	0.166	-0.119	0.124	0.173	

5.3 Comparison with alternative estimators

We compare the performance of the PCML estimator for model (18) with two alternative approaches. The first, denoted by W, is the correlated random-effects approach based on the proposal by Wooldridge (2005) for nonlinear dynamic panel data models, where the individual unobserved heterogeneity is assumed to be normally distributed and initial conditions are handled by specifying the distribution of c_i conditional on the initial value of $\boldsymbol{y}_i$. In Wooldridge (2005) a general formulation for this conditional distribution is proposed, where the individual random effects are allowed to depend on linear combinations of time-averages of strictly exogenous covariates (Mundlak, 1978). We specify the following conditional distribution of c_i

$$c_i|y_{i1} \sim y_{i1}\alpha + \bar{v}_i\pi + c_i^*, \qquad c_i^* \sim N(0, \sigma_c^2), \quad i = 1, \ldots, n.$$

where $\bar{v}_i = (1/T)\sum_{t=1}^{T} v_{it}$. It is worth noting that, in this case, the ML estimator of the model parameters is consistent if c_i^* is independent of the possibly predetermined covariate x_{it}. Therefore, we generate samples where c_i in (30) is distributed as a normal random variable with zero mean, unit variance, and $\varpi = 0$, in order to avoid the misspecification of the random effects. Nevertheless we also compare the ML and PCML estimator in the scenario where the individual intercepts are generated as in (30).

The second is the so-called infeasible logit estimator (Honoré and Kyriazidou, 2000) denoted by INF, where the generated individual intercepts are used as an additional covariate and the model parameters are then estimated by ML based on the pooled logit model formulation. The purpose is to compare the PCML estimator with a benchmark that is not sensitive to substantial deviations from the approximating model (20).

Tables 7 and 8 summarize the results of the simulation study, that we limit to the scenarios with $\beta = -1$, $\gamma = -1$, and $\eta = 0, -1$. Table 7 contains the results based on the samples generated with individual effects independent of the model covariate. The biases for β obtained by PCML and W are similar to those obtained by the infeasible logit, especially with $T = 8$, and the RMSE and MAE attain the same order of magnitude to those of INF with $n = 1000$ and $T = 8$. With $\eta = -1$, W shows a small bias for β and values of RMSE and MAE similar to the PCML estimator. As for γ, the bias of W increases with both values of η. This result is likely due to the fact that the actual number of time occasions exploited by the ML estimator is too small for W to deliver a negligible bias, for which at least 8 occasions are required (Akay, 2012). As expected, though, W exhibits rather large biases when the individual intercepts are generated as in (30) with $\varpi = 0.25$ (see Table 8), which makes the PCML a more attractive alternative since this is a scenario that is more likely to occur in practice.

Table 7: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $c_i \sim N(0,1)$, $\varpi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML	0.003	0.204	0.024	0.133	0.063	0.011	0.454	-0.006	0.317	0.051	0.059
W	-0.013	0.131	-0.003	0.090	0.054	0.030	0.279	0.031	0.199	0.045	0.046
INF	-0.012	0.094	-0.013	0.063	0.045	0.013	0.102	0.012	0.066	0.051	0.053
					$n = 500$, $T = 8$						
PCML	-0.003	0.064	0.002	0.045	0.049	0.005	0.121	0.004	0.080	0.047	0.046
W	-0.002	0.060	-0.000	0.041	0.045	0.005	0.112	0.005	0.076	0.045	0.040
INF	-0.002	0.055	-0.001	0.037	0.055	0.005	0.056	0.003	0.038	0.047	0.043
					$n = 1000$, $T = 4$						
PCML	0.022	0.138	0.032	0.092	0.052	-0.023	0.294	-0.031	0.197	0.045	0.060
W	-0.003	0.095	0.004	0.065	0.060	0.025	0.203	0.034	0.138	0.061	0.051
INF	-0.005	0.070	-0.003	0.047	0.064	0.007	0.069	0.007	0.047	0.042	0.037
					$n = 1000$, $T = 8$						
PCML	-0.000	0.046	-0.000	0.032	0.056	-0.003	0.083	-0.003	0.056	0.040	0.064
W	0.002	0.042	0.003	0.028	0.062	-0.003	0.079	-0.002	0.052	0.042	0.056
INF	0.002	0.038	0.001	0.024	0.057	0.002	0.040	0.002	0.028	0.041	0.044
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML	-0.010	0.279	0.020	0.173	0.058	0.023	0.559	-0.004	0.343	0.049	0.993
W	-0.059	0.187	-0.035	0.118	0.044	0.061	0.370	0.072	0.254	0.059	1.000
INF	-0.014	0.113	-0.006	0.074	0.054	0.007	0.120	0.003	0.078	0.045	1.000
					$n = 500$, $T = 8$						
PCML	0.025	0.085	0.027	0.057	0.061	-0.030	0.162	-0.029	0.111	0.050	1.000
W	-0.051	0.090	-0.049	0.061	0.089	0.104	0.180	0.103	0.124	0.118	1.000
INF	-0.004	0.065	-0.001	0.043	0.047	0.002	0.070	-0.001	0.047	0.051	1.000
					$n = 1000$, $T = 4$						
PCML	0.016	0.182	0.028	0.128	0.048	-0.025	0.379	-0.028	0.242	0.044	1.000
W	-0.041	0.134	-0.036	0.087	0.059	0.063	0.268	0.062	0.186	0.060	1.000
INF	-0.010	0.081	-0.009	0.055	0.055	0.011	0.085	0.011	0.057	0.044	1.000
					$n = 1000$, $T = 8$						
PCML	0.025	0.064	0.026	0.044	0.077	-0.038	0.121	-0.038	0.080	0.071	1.000
W	-0.050	0.073	-0.050	0.053	0.148	0.096	0.143	0.098	0.105	0.165	1.000
INF	-0.000	0.046	0.000	0.030	0.052	0.003	0.050	0.001	0.033	0.061	1.000

Table 8: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $c_i = (1/T)\sum_{t=1}^{4} x_{it}^*$, $\varpi = 0.5$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML	-0.002	0.194	0.018	0.123	0.053	-0.011	0.432	-0.020	0.290	0.055	0.057
W	0.162	0.205	0.167	0.169	0.311	-0.205	0.362	-0.197	0.245	0.085	0.663
INF	-0.011	0.098	-0.008	0.066	0.058	0.016	0.094	0.015	0.064	0.049	0.065
					$n = 500$, $T = 8$						
PCML	-0.011	0.065	-0.009	0.045	0.052	0.005	0.118	0.003	0.078	0.041	0.054
W	0.056	0.082	0.058	0.061	0.183	-0.061	0.125	-0.064	0.089	0.067	0.277
INF	-0.005	0.054	-0.006	0.036	0.051	0.005	0.056	0.005	0.038	0.055	0.050
					$n = 1000$, $T = 4$						
PCML	0.028	0.132	0.038	0.094	0.058	-0.010	0.305	-0.009	0.206	0.056	0.051
W	0.173	0.194	0.176	0.176	0.547	-0.196	0.289	-0.193	0.212	0.148	0.915
INF	-0.005	0.068	-0.006	0.045	0.059	0.006	0.067	0.009	0.045	0.058	0.053
					$n = 1000$, $T = 8$						
PCML	-0.003	0.043	-0.004	0.028	0.047	0.006	0.088	0.004	0.057	0.058	0.045
W	0.063	0.074	0.062	0.062	0.325	-0.060	0.100	-0.060	0.070	0.124	0.488
INF	-0.000	0.037	-0.000	0.025	0.037	0.000	0.039	-0.000	0.026	0.051	0.037
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML	0.002	0.276	0.021	0.177	0.058	0.003	0.534	-0.023	0.356	0.039	0.996
W	0.057	0.200	0.072	0.143	0.101	0.007	0.416	0.037	0.293	0.060	1.000
INF	-0.068	0.130	-0.068	0.083	0.074	0.229	0.255	0.226	0.226	0.517	1.000
					$n = 500$, $T = 8$						
PCML	0.023	0.086	0.021	0.060	0.059	-0.030	0.163	-0.033	0.114	0.055	1.000
W	-0.020	0.079	-0.016	0.055	0.062	0.060	0.158	0.058	0.107	0.073	1.000
INF	-0.016	0.069	-0.016	0.046	0.056	0.117	0.135	0.116	0.116	0.399	1.000
					$n = 1000$, $T = 4$						
PCML	0.023	0.183	0.040	0.122	0.050	-0.022	0.370	-0.028	0.245	0.045	1.000
W	0.072	0.152	0.075	0.107	0.129	0.007	0.298	0.011	0.203	0.066	1.000
INF	-0.063	0.102	-0.060	0.070	0.114	0.222	0.236	0.220	0.220	0.814	1.000
					$n = 1000$, $T = 8$						
PCML	0.024	0.062	0.025	0.042	0.075	-0.029	0.115	-0.032	0.084	0.059	1.000
W	-0.021	0.057	-0.022	0.039	0.056	0.057	0.116	0.058	0.081	0.076	1.000
INF	-0.018	0.050	-0.017	0.033	0.061	0.113	0.123	0.113	0.113	0.669	1.000

6 Conclusions

In this paper, we propose a novel model formulation for dynamic binary panel data models that accounts for feedback effects from the past of the outcome variable on the present value of 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 for microeconomic applications. Our formulation is based on the equivalence between Granger's definition of noncausality and a modification of the Sims' strict exogeneity assumption for nonlinear panel data models, introduced by Chamberlain (1982) and for which we provide a more general theorem.

Under the logit model, the proposed model formulation yields three 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 allows, in practice, for the inclusion of a large number of predetermined covariates, discrete or continuous; (*iii*) its parameters can be estimated within a fixed-effects approach by a PCML, thereby allowing for an arbitrary dependence structure between the model covariates and the individual permanent unobserved heterogeneity.

From our simulation results, it emerges that PCML provides consistent estimation of the regression and state dependence parameters in presence of substantial departures from noncausality and that the bias is negligible even when the conditions for the exact logit model formulation are violated. Furthermore, we show that the alternative random-effects ML estimator based on Wooldridge (2005) for the model here proposed exhibits comparable finite-sample properties, provided the dependence between the predetermined covariate and the unobserved heterogeneity is reliably accounted for.

Finally, the logit model here proposed is fairly easy to estimate using available software. The PCML estimator of the proposed model can be implemented using the package `cquad` (Bartolucci and Pigini, 2016), whereas any routine for the random-effects logit model can be used for correlated-random effects ML estimator.

References

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics*, 27(7):1189–1204.

Alessie, R., Hochguertel, S., and van Soest, A. (2004). Ownership of Stocks and Mutual Funds: A Panel Data Analysis. *The Review of Economics and Statistics*, 86(3):783–796.

Anderson, T. W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375):598–606.

Arellano, M. and 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(2):277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.

Arellano, M. and Carrasco, R. (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics*, 115(1):125–157.

Arulampalam, W. (2002). State dependence in unemployment incidence: evidence for british men revisited. Technical report, IZA Discussion paper series.

Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons.

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

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

Bartolucci, F. and Pigini, C. (2016). `cquad`: An R and Stata package for conditional maximum likelihood estimation of dynamic binary panel data models. *Journal of Statistical Software*, In press.

Bettin, G. and Lucchetti, R. (2016). Steady streams and sudden bursts: persistence patterns in remittance decisions. *Journal of Population Economics*, 29(1):263–292.

Biewen, M. (2009). Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. *Journal of Applied Econometrics*, 24(7):1095–1116.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.

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

Cappellari, L. and Jenkins, S. P. (2004). Modelling low income transitions. *Journal of Applied Econometrics*, 19(5):593–610.

Carrasco, R. (2001). Binary choice with binary endogenous regressors in panel data. *Journal of Business & Economic Statistics*, 19(4):385–394.

Carro, J. M. and Traferri, A. (2012). State dependence and heterogeneity in health using a bias-corrected fixed-effects estimator. *Journal of Applied Econometrics*, 29(2):181–207.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47(1):225–238.

Chamberlain, G. (1982). The general equivalence of granger and sims causality. *Econometrica: Journal of the Econometric Society*, 50(3):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. and Singer, B., editors, *Longitudinal Analysis of Labor Market Data*. Cambridge University Press: Cambridge.

Contoyannis, P., Jones, A. M., and Rice, N. (2004). Simulation-based inference in dynamic panel probit models: an application to health. *Empirical Economics*, 29(1):49–77.

Cox, D. (1972). The analysis of multivariate binary data. *Applied statistics*, 21(2):113–120.

Florens, J.-P. and Mouchart, M. (1982). A note on noncausality. *Econometrica: Journal of the Econometric Society*, 50(3):583–591.

Giarda, E. (2013). Persistency of financial distress amongst italian households: Evidence from dynamic models for binary panel data. *Journal of Banking & Finance*, 37(9):3425 – 3434.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3):424–438.

Halliday, T. J. (2008). Heterogeneity, state dependence and health. *The Econometrics Journal*, 11(3):499–516.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 50(4):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).

Heckman, J. J. and Borjas, G. J. (1980). Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica*, 47(187):pp. 247–283.

Heiss, F. (2011). Dynamics of self-rated health and selective mortality. *Empirical economics*, 40(1):119–140.

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

Honoré, B. E. and Lewbel, A. (2002). Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica*, 70(5):2053–2063.

Hsiao, C. (2005). *Analysis of Panel Data*. Cambridge University Press, New York, 2nd edition.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67(6):1255–1294.

Keane, M. P. and Sauer, R. M. (2009). Classification error in dynamic discrete choice models: Implications for female labor supply behavior. *Econometrica*, 77(3):975–991.

Michaud, P.-C. and Tatsiramos, K. (2011). Fertility and female employment dynamics in europe: the effect of using alternative econometric modeling assumptions. *Journal of Applied Econometrics*, 26(4):641–668.

Mosconi, R. and Seri, R. (2006). Non-causality in bivariate binary time series. *Journal of Econometrics*, 132(2):379–407.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

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

Sims, C. A. (1972). Money, income, and causality. *The American Economic Review*, 62(4):540–552.

Stewart, M. B. (2007). The interrelated dynamics of unemployment and low-wage employment. *Journal of Applied Econometrics*, 22(3):511–531.

Wooldridge, J. M. (2000). A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. *Economics Letters*, 68(3):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(1):39–54.

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

Wunder, C. and Riphahn, R. T. (2014). The dynamics of welfare entry and exit amongst natives and immigrants. *Oxford Economic Papers*, 66(2):580–604.