



Recursive Computation of the Conditional Probability Function of the Quadratic Exponential Model for Binary Panel Data

Francesco Bartolucci¹ · Francesco Valentini² · Claudia Pigini²

Accepted: 19 October 2021 / Published online: 18 November 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

We propose a general recursive algorithm for the computation of the conditional probability function of the quadratic exponential model for binary panel data given the total of the responses, which is a sufficient statistic for the individual intercept parameter. This recursion permits to implement conditional and pseudo-conditional maximum likelihood estimators of the parameters of this model, and related models such as the dynamic logit model, even when one or more statistical units are observed at many occasions. In this way we solve a typical problem in dealing with distributions with a complex normalizing constant. The advantage in terms of computational load with respect to standard techniques is assessed by simulation and illustrated by an application based on a popular dataset about brand loyalty.

Keywords Clustered data · Conditional maximum likelihood estimation · Longitudinal data · Newton-Raphson algorithm

1 Introduction

Binary choice models for longitudinal binary data play an important role in the recent statistical and econometric literature (Diggle et al., 2002; Hsiao, 2014). The simplest models of this type are based on a logit or probit transformation of the probability of the binary response being equal to 1, which is assumed to depend on individual intercepts, a linear function of the covariates, and, possibly, the lagged response variable. These models can be formulated following a random- or a fixed-

✉ Francesco Bartolucci
francesco.bartolucci@unipg.it

¹ Department of Economics, University of Perugia, Via A. Pascoli, 20, 06123 Perugia, Italy

² Department of Economics and Social Sciences, Marche Polytechnic University, P.le Martelli, 8, 60121 Ancona, Italy

effects approach, according to whether individual intercepts are treated as random variables or fixed parameters to be estimated. The main feature of the latter approach is that it does not require distributional assumptions for the individual-specific parameters, with advantages in terms of robustness of the conclusions that may be reached on the basis of the model.

The Maximum Likelihood (ML) estimator for the parameters of a binary fixed-effects model for panel data is however inconsistent because of the incidental parameters problem (Neyman & Scott, 1948; Lancaster, 2000). The incidental parameters are the individual intercepts and a relevant bias in their estimation arises whenever sample units are observed over a small number of time occasions T , so that the ML estimator is based only on a few observations per unit. Moreover, this bias spreads to the estimator of the regression parameters, as they are not informationally orthogonal to the incidental parameters, and is of order $O(T^{-1})$.

The literature proposes two different approaches to overcome the methodological issues arising within the ML estimation of fixed-effects binary choice models. A first approach is focused on reducing the order of the bias of the ML estimator to $O(T^{-2})$, by means of bias corrections of the parameter estimator (Hahn & Newey, 2004; Fernández-Val, 2009; Hahn & Kuersteiner, 2011; Dhaene & Jochmans, 2015) or by using a modified likelihood (Bester & Hansen, 2009; Arellano & Hahn, 2016; Bartolucci et al., 2016) or score function (Carro, 2007). This strategy is very general and can be adapted to both probit and logit as well as to other nonlinear models. It can also accommodate dynamic formulations that include the lagged response variable among the covariates, which is sensible for a variety of economic applications. However, these estimators exhibit poor finite-sample performance when T is small, a situation that is likely to occur in practice with rotated surveys or even with long panels, if these are highly unbalanced.

The second approach is represented by conditional inference (Andersen, 1970, 1971), which is based on conditioning the joint probability of the response configuration provided by each subject on a sufficient statistic for the specific intercept, thereby eliminating the incidental parameters problem. Differently from the bias-corrected ones, the resulting Conditional Maximum Likelihood (CML) estimators are consistent as the sample size, n , goes to infinity with T fixed. The viability of this strategy is however tied to the existence of reasonable sufficient statistics for the incidental parameters for the specific model formulations.

With binary panel data, CML estimation has been applied to the static logit (Chamberlain, 1980; Arellano & Honoré, 2001; Wooldridge, 2010) and Quadratic Exponential (QE) model (Cox, 1972a; Bartolucci & Nigro, 2010). Traditionally, CML for binary clustered data has also been adopted in psychometrics to obtain consistent item estimates in the popular Rasch model (1960, 1961) and related latent class models (Lindsay et al., 1991; Agresti, 2003). Conditional inference has also been applied to generalized additive mixed models (Zhang & Davidian, 2004). More recently, simplified versions based on pairwise conditional likelihood approximations have been used to estimate the homophily parameters of network formation models by Charbonneau (2017) and Jochmans (2018), while Graham

(2017) advocates the use of CML to estimate the parameters of the *tetrad* logit for undirected networks; see also de Paula (2020) for a recent review.

Sufficient statistics for the individual intercepts are not available in the case of the Dynamic Logit (DL) model, for which a weighted CML estimator has been derived by Honoré & Kyriazidou (2000) and a Pseudo-CML (PCML) estimator based on an approximating QE model has been proposed by Bartolucci & Nigro (2012); see also Bartolucci & Pennoni (2007).

From a practitioner perspective, CML estimators are appealing for economic applications based on small- T panel data, as well as on long, but highly unbalanced, longitudinal surveys. However, CML estimators require the maximization of peculiar likelihood functions, whose computation in the standard way implies an excessive burden when T is large, thus limiting the applicability of these techniques that may become infeasible in certain situations. One way to overcome these computational issues is to exploit recursive algorithms as we here propose.

Several works deal with the recursive computation of the log-likelihood function for a variety of models in the conditional inference framework. The seminal work of Howard (1972) shows that the conditional likelihood function of the model proposed by Cox (1972b) is a symmetric function and it can be written recursively. Another popular application of recursive algorithms is relative to the analysis of epidemiological stratified case-control studies. Smith et al. (1981) propose an algorithm for the CML estimation of the coefficients of a logistic regression and, since this procedure becomes computationally challenging for large strata, Krailo and Pike (1984) derive a recursive computation of the log-likelihood and its derivatives. Moreover, Levin (1987) extends this result to models for multinomial outcomes. Other recursive solutions come from Item Response Theory (Hambleton & Swaminathan, 1986; Bartolucci et al., 2015), where Andersen (1972) and Gustafsson (1980) show a recursive structure for the conditional estimating equations of the Rasch model (Rasch 1960).

In this paper, we propose a general recursive algorithm to compute the conditional likelihood of a series of models for binary data in an efficient way, so that the CML approach may be easily applied even when a large number of observations is available for a few sample units. The main contribution is represented by the generalization of the result by Krailo and Pike (1984) to the QE model and its extensions to accommodate the dynamic formulations recently proposed by Bartolucci & Nigro (2010), Bartolucci & Nigro (2012), and Bartolucci et al. (2018), whose computation would otherwise be unfeasible with a large time dimension. Furthermore, we implement the proposed recursive algorithm in the `cquad` package¹, which provides software routines for the CML estimation of QE model and the related aforementioned contributions (Bartolucci & Pigini, 2017). A Monte Carlo simulation shows how the proposed algorithm avoids the computational burden of the QE model for large- T datasets. This broadens the applicability of conditional inference for dynamic models in many economic applications. As an example, this work includes an empirical application concerning

¹ Available for the software R (<https://cran.r-project.org/package=cquad>) and Stata (<https://ideas.repec.org/c/boc/bocode/s458852.html>, <https://github.com/fravale/cquadr>).

brand loyalty based on real data on yogurt purchases, which is a popular example in economics (Chintagunta et al., 2001).

The present work is organized as follows. Section 2 illustrates the main theoretical aspects of the models and the conditional estimation framework. Section 3 shows the proposed recursive algorithm for the QE model and its extensions. The performance of the algorithm in terms of computational time is evaluated by the Monte Carlo simulation in Sect. 4 and the analysis of brand loyalty data is illustrated in Sect. 5. Finally, Sect. 6 provides the main conclusions.

2 Model Assumptions

This section first introduces the main theoretical aspects of the fixed-effects logit and QE model. Then, the CML estimator for both models is presented.

2.1 Static Logit Model

Consider a sample of n individuals, where each individual i , $i = 1, \dots, n$, is observed at T_i time occasions, with T_i allowed to vary across individuals in order to account for unbalanced panel structures.

Let y_{it} be the binary response for subject i at occasion t and \mathbf{x}_{it} a column vector of covariates, where $t = 1, \dots, T_i$, with $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT_i})$ being the matrix collecting all such covariates. Consider the logit formulation

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

where α_i is the individual intercept, capturing the time-constant subject-specific unobserved heterogeneity, and $\boldsymbol{\beta}$ collects the common slope parameters.

As mentioned in Sect. 1, the conditional inference approach (Andersen, 1970) has been applied to derive the conditional likelihood for the static logit model by Chamberlain (1980). Consider the joint probability of the individual response configuration $\mathbf{y}_i = (y_{i1}, \dots, y_{iT_i})'$ that, under the logit model, is

$$p(\mathbf{y}_i|\alpha_i, \mathbf{X}_i) = \prod_{t=1}^{T_i} p(y_{it}|\alpha_i, \mathbf{X}_i) = \frac{\exp[y_{i+}\alpha_i + (\sum_{t=1}^{T_i} y_{it}\mathbf{x}_{it})'\boldsymbol{\beta}]}{\prod_{t=1}^{T_i} [1 + \exp(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta})]}, \tag{2}$$

where $y_{i+} = \sum_{t=1}^{T_i} y_{it}$ is the *total score*, which can be proved to be a sufficient statistic for α_i . In this regard, consider the probability of observing a given total score for the logit model so that, conditional on α_i and \mathbf{X}_i , we have

$$p(y_{i+}|\alpha_i, \mathbf{X}_i) = \frac{\sum_{\mathbf{z}:\mathbf{z}_+=y_{i+}} \exp(z_+\alpha_i) \exp\left[\left(\sum_{t=1}^{T_i} z_t\mathbf{x}_{it}\right)'\boldsymbol{\beta}\right]}{\prod_{t=1}^{T_i} [1 + \exp(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta})]}, \tag{3}$$

where the numerator corresponds to the sum of all possible binary response vectors

$\mathbf{z} = (z_1, \dots, z_{T_i})'$ such that $z_+ = \sum_{t=1}^{T_i} z_t$ equals y_{i+} . The probability of y_i conditional on α_i, \mathbf{X}_i , and y_{i+} is the ratio between expressions (2) and (3), so that

$$\begin{aligned}
 p(\mathbf{y}_i | \alpha_i, \mathbf{X}_i, y_{i+}) &= \frac{\exp(y_{i+} \alpha_i) \exp\left[\left(\sum_{t=1}^{T_i} y_{it} \mathbf{x}_{it}\right)' \boldsymbol{\beta}\right]}{\prod_{t=1}^{T_i} [1 + \exp(\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta})]} \\
 &\quad \times \frac{\prod_{t=1}^{T_i} [1 + \exp(\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta})]}{\sum_{\mathbf{z}: z_+ = y_{i+}} \exp(z_+ \alpha_i) \exp\left[\left(\sum_{t=1}^{T_i} z_t \mathbf{x}_{it}\right)' \boldsymbol{\beta}\right]} \tag{4} \\
 &= \frac{q_1(\mathbf{X}_i, y_i)}{\sum_{\mathbf{z}: z_+ = y_{i+}} q_1(\mathbf{X}_i, \mathbf{z})} = p(y_i | \mathbf{X}_i, y_{i+}),
 \end{aligned}$$

where

$$q_1(\mathbf{X}_i, y_i) = \exp\left(\sum_{t=1}^{T_i} y_{it} \mathbf{x}'_{it} \boldsymbol{\beta}\right). \tag{5}$$

Notice that the conditional probability $p(y_i | \mathbf{X}_i, y_{i+})$ no longer depends on the individual intercept.

On the basis of the previous results we obtain the conditional log-likelihood function

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbb{1}\{0 < y_{i+} < T_i\} \log p(y_i | \mathbf{X}_i, y_{i+}),$$

with $\mathbb{1}\{\cdot\}$ being the indicator function, which is maximized by a Newton-Raphson algorithm in order to obtain a CML estimate.

Starting from an initial guess for the parameter vector, denoted by $\boldsymbol{\beta}^{(0)}$, at step h the algorithm updates the parameter vector obtained at the previous step through the usual rule

$$\boldsymbol{\beta}^{(h)} = \boldsymbol{\beta}^{(h-1)} - \left[\frac{\partial^2 \ell(\boldsymbol{\beta}^{(h-1)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right]^{-1} \frac{\partial \ell(\boldsymbol{\beta}^{(h-1)})}{\partial \boldsymbol{\beta}},$$

where

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \mathbb{1}\{0 < y_{i+} < T_i\} \frac{\partial \log p(y_i | \mathbf{X}_i, y_{i+})}{\partial \boldsymbol{\beta}}, \tag{6}$$

$$\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^n \mathbb{1}\{0 < y_{i+} < T_i\} \frac{\partial^2 \log p(y_i | \mathbf{X}_i, y_{i+})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}. \tag{7}$$

This is repeated until convergence obtaining the CML estimate $\hat{\boldsymbol{\beta}}$. Note also that minus the Hessian matrix obtained at convergence may be used to estimate the variance-covariance matrix for $\hat{\boldsymbol{\beta}}$, and then its standard errors, in the usual way.

It is important noting that computation of $p(\mathbf{y}_i | \mathbf{X}_i, y_{i+})$ is feasible only with moderate values of $T = \max_i T_i$ because expression (4) involves a sum over the binary vectors \mathbf{z} with a fixed total score, the number of which quickly grows in T , as will be detailed in Sect. 4. This is a typical problem of computation of a normalizing constant that occurs in many statistical contexts as, just to mention one of the many relevant cases, in the models for spatial data described in the seminal paper of Besag (1974).

2.2 Dynamic Models

The DL model is a straightforward extension of the static logit model in (1) as it includes as a regressor the lagged dependent variable, $y_{i,t-1}$, along with the exogenous covariates and the individual intercept. The probability of the response variable under the DL model is

$$p(y_{it} | \alpha_i, \mathbf{X}_i, y_{i0}, \dots, y_{i,t-1}) = \frac{\exp[y_{it}(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma)]}{1 + \exp(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma)}, \tag{8}$$

where the parameter γ measures the so-called state dependence (Heckman, 1981), which represents how much the experience of a certain event affects the probability of experiencing the same event in the future. Here the conditioning set also includes the past values of the response variable, with the initial observation y_{i0} assumed to be known.

The probability of the response configuration \mathbf{y}_i under the DL model is

$$p(\mathbf{y}_i | \alpha_i, \mathbf{X}_i, y_{i0}) = \frac{\exp[y_{i+}\alpha_i + (\sum_{t=1}^{T_i} y_{it}\mathbf{x}_{it})'\boldsymbol{\beta} + y_{i*}\gamma]}{\prod_{t=1}^{T_i} [1 + \exp(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma)]}, \tag{9}$$

where $y_{i*} = \sum_{t=1}^{T_i} y_{i,t-1}y_{it}$. Under the DL model, the total score is no longer a sufficient statistic for the individual parameter α_i . Therefore, Chamberlain (1993) derives a CML estimator for the DL model only for the case with no covariates and when $T = 3$. Moreover, Honoré & Kyriazidou (2000) propose CML estimation of the DL model with explanatory variables by exploiting a weighted conditional log-likelihood. However, their strategy rules out the use of trending regressors and dummies; the resulting estimator has a rate of convergence that is slower than \sqrt{n} .

Among the dynamic formulations for binary panel data, the QE model closely resembles the DL model, as it allows us to include individual specific effects and the lagged dependent variable among the covariates. This model directly formulates the joint probability of \mathbf{y}_i as

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

Here δ_i denotes the individual specific effects, $\boldsymbol{\eta}_1$ is the vector of regression parameters, and $\boldsymbol{\psi}$ measures the true state dependence. In this formulation, $\boldsymbol{\phi}$ and $\boldsymbol{\eta}_2$

are considered nuisance parameters. The denominator is based on the sum extended to all the possible binary response vectors \mathbf{z} , while $z_{i*} = y_{i0}z_1 + \sum_{t=2}^{T_i} z_{t-1}z_t$.

Differently from the DL model, in the QE model the total score y_{i+} is a sufficient statistic for the individual intercept δ_i , so that the conditional likelihood function is based on the individual contributions

$$p(\mathbf{y}_i | \mathbf{X}_i, y_{i0}, y_{i+}) = \frac{\exp \left[\left(\sum_{t=1}^{T_i} y_{it} \mathbf{x}_{it}' \right) \boldsymbol{\eta}_1 + y_{iT_i} \left(\boldsymbol{\phi} + \mathbf{x}'_{iT_i} \boldsymbol{\eta}_2 \right) + y_{i*} \psi \right]}{\sum_{\mathbf{z}: z_+ = y_{i+}} \exp \left[\left(\sum_{t=1}^{T_i} z_{it} \mathbf{x}_{it}' \right) \boldsymbol{\eta}_1 + z_{T_i} \left(\boldsymbol{\phi} + \mathbf{x}'_{iT_i} \boldsymbol{\eta}_2 \right) + z_{i*} \psi \right]}. \quad (10)$$

The CML estimator can easily be obtained by Newton-Raphson algorithm. Even in this case, though, estimation is unfeasible when T is large for some reasons illustrated above for the static logit model.

2.3 Simplified, Modified, and Approximating QE Models

We now focus on three models deriving from the QE model in Eq. (10). These are: (i) the simplified QE model outlined in Bartolucci & Pignini (2017); (ii) the modified QE model put forward in Bartolucci et al. (2018); and (iii) the approximating QE version for the DL model proposed by Bartolucci & Nigro (2012). A common feature of these models is that the conditional probability for an individual response configuration can be written as

$$p(\mathbf{y}_i | \mathbf{X}_i, y_{i0}, y_{i+}) = \frac{q_j(\mathbf{X}_i, \mathbf{y}_i)}{\sum_{\mathbf{z}: z_+ = y_{i+}} q_j(\mathbf{X}_i, \mathbf{z})}, \quad j = 2, 3, 4, \quad (11)$$

respectively.

The simplified QE model assumes that the regression parameters are the same for all time occasions, implying a reduction of the set of parameters in Eq. (10) from $(\boldsymbol{\eta}_1, \boldsymbol{\phi}, \boldsymbol{\eta}_2, \psi)$ to $(\boldsymbol{\eta}, \psi)$, so that the probability in Eq. (11) can be written as a function of

$$q_2(\mathbf{X}_i, \mathbf{y}_i) = \exp \left(\sum_{t=1}^{T_i} y_{it} \mathbf{x}'_{it} \boldsymbol{\eta} + y_{i*} \psi \right). \quad (12)$$

The modified QE model closely resembles the simplified QE version. The main difference is that, in this case, the association rule between the outcome variable and its lag takes into account the pairs of consecutive observations that are equal, regardless of their value being 0 or 1. This formulation is such that a more powerful test for $H_0 : \psi = 0$ (absence of state dependence) can be obtained with respect to the QE and simplified QE models.

Under the modified QE model, the conditional probability of \mathbf{y}_i given \mathbf{X}_i and y_{i0} can be written as a function of

$$q_3(\mathbf{X}_i, \mathbf{y}_i) = \exp\left(\sum_{t=1}^{T_i} y_{it} \mathbf{x}'_{it} \boldsymbol{\eta} + \tilde{y}_{i*} \psi\right), \quad (13)$$

where $\tilde{y}_{i*} = \sum_{t=1}^{T_i} \mathbb{1}\{y_{it} = y_{i,t-1}\}$ is the number of consecutive observations equal each other, being other 0 or 1. The corresponding denominator in (11) is the sum over all the \mathbf{z} such that $z_+ = y_{i+}$ of $q_3(\mathbf{X}_i, \mathbf{z})$, which will be a function of $\tilde{z}_{i*} = \mathbb{1}\{z_{i1} = y_{i0}\} + \sum_{t>1} \mathbb{1}\{z_{it} = z_{i,t-1}\}$.

Finally, Bartolucci & Nigro (2012) show that the QE model can serve as an approximation of the DL model. The approximating formulation is based on a first-order Taylor expansion of (9) around $\alpha = \bar{\alpha}$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, and $\gamma = 0$. The main advantage is that the approximating QE model admits the total scores as sufficient statistics for the individual intercepts. For this model, the conditional probability of \mathbf{y}_i is a function of

$$q_4(\mathbf{X}_i, \mathbf{y}_i) = \exp\left(\sum_{t=1}^{T_i} y_{it} \mathbf{x}'_{it} \boldsymbol{\beta} - \sum_{t=1}^{T_i} \bar{r}_{it} y_{i,t-1} \gamma + \sum_{t=1}^{T_i} y_{i*} \gamma\right), \quad (14)$$

where $\bar{r}_{it} = \exp(\bar{\alpha}_i + \mathbf{x}'_{it} \bar{\boldsymbol{\beta}}) / [1 + \exp(\bar{\alpha}_i + \mathbf{x}'_{it} \bar{\boldsymbol{\beta}})]$.

The model parameters are estimated by PCML on the basis of two steps: first, \bar{r}_{it} is computed using a CML estimate for $\bar{\boldsymbol{\beta}}$ and $\bar{\alpha}_i$ equal to its ML estimate under the static logit model; second, the estimate of $(\boldsymbol{\beta}', \gamma)'$ is obtained by maximizing the conditional log-likelihood based on (14).

Overall, all models here discussed, included the static logit model, may be expressed in the form (11) with different definitions of function $q_j(\mathbf{X}_i, \mathbf{y}_i)$, $j = 1, \dots, 4$, that are defined in Eqs. (5), (12), (13), and (14), respectively.

3 Recursive Computation of the Conditional Likelihood Functions

A common shortcoming of the aforementioned models is the computational burden in evaluating the normalizing constants $\sum_{\mathbf{z}:z_+=y_{i+}} q_j(\mathbf{X}_i, \mathbf{z})$, $j = 1, \dots, 4$. This section first introduces the recursive algorithm by Krailo & Pike (1984) for the static logit model. Then it shows the proposed extension for evaluating the normalizing constants for the QE formulations, namely the simplified, modified and approximating QE models.

3.1 Recursion for the Static Logit Model

Consider the denominator of the conditional probability in (4), which is given by

$$\sum_{\mathbf{z}:z_+=y_{i+}} q_1(\mathbf{X}_i, \mathbf{z}), \quad (15)$$

where $q_1(\mathbf{X}_i, \mathbf{z})$ is defined in (5).

From the computational point of view, the sum above represents a limitation to the applicability of the CML estimator, especially with large T , as will be discussed

is Sect. 4. However, as noted by Howard (1972), the computation of the sum in Eq. (15) can be performed recursively. Define now for individual i function $f_{t,s}(\cdot)$, indexed by the time dimension $t = 1, \dots, T_i$ and by the total score denoted $s = 0, \dots, y_{i+}$, as

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

where now $\mathbf{z} = (z_1, \dots, z_t)'$ such that $\sum_{u=1}^t z_u = s$ and $\phi_u = \mathbf{x}'_{iu} \boldsymbol{\beta}$. Further, define $\boldsymbol{\phi} = (\phi_1, \dots, \phi_t)'$. Note that $f_{t,s}(\boldsymbol{\phi})$ corresponds to the quantity defined in (15) for $t = T_i$ and $s = y_{i+}$, that is,

$$\sum_{\mathbf{z}:z_+=y_{i+}} q_1(\mathbf{X}_i, \mathbf{z}) = f_{T_i, y_{i+}}(\boldsymbol{\phi}). \tag{16}$$

By exploiting the symmetric function properties of $f_{t,s}(\boldsymbol{\phi})$, it can be computed recursively in this way:

1. for $t = 1$, let $f_{1,0}(\boldsymbol{\phi}) = 1, f_{1,1}(\boldsymbol{\phi}) = \exp(\phi_1)$, and $0 f_{2,2} = f_{1,2} + f_{1,1}$ otherwise;
2. for $t = 2, \dots, T_i$, compute

$$f_{t,s}(\boldsymbol{\phi}) = f_{t-1,s}(\boldsymbol{\phi}) + f_{t-1,s-1}(\boldsymbol{\phi}) \exp(\phi_t), \quad s = 0, \dots, t. \tag{17}$$

The recursive structure can be exploited for the computation the first- and the second-derivatives of this function as well. In turn by exploiting (16), these derivatives are used to compute the score vector and the Hessian matrix in (6) and (7), which are involved in the Newton-Raphson algorithm for the maximization of this function. For the first and second derivatives and for $h, j = 1, \dots, T_i, s = 0, \dots, t$, and $t = 1, \dots, T_i$, we use the notation

$$f_{t,s}^{(h)}(\boldsymbol{\phi}) = \frac{\partial f_{t,s}(\boldsymbol{\phi})}{\partial \phi_h}, \quad \text{and} \quad f_{t,s}^{(h,j)}(\boldsymbol{\phi}) = \frac{\partial^2 f_{t,s}(\boldsymbol{\phi})}{\partial \phi_h \partial \phi_j},$$

respectively. How to compute these derivatives by a recursion for the static logit, as well as for the other QE models, is described in “Appendix”.

3.2 QE Models

The methodology outlined in the previous section for the static logit model cannot be applied directly to the QE models illustrated in Sect. 2.3. In the following, we propose recursive algorithms to deal with these models.

Similarly to the static logit model, given the time dimension, the total score, and the initial observation, the sum at the denominator of Eq. (11) may be expressed as a function of type

$$g_{t,a,s}(\boldsymbol{\phi}, \psi) = g_{t,a,s,v=0}(\boldsymbol{\phi}, \psi) + g_{t,a,s,v=1}(\boldsymbol{\phi}, \psi), \quad (18)$$

where the index v corresponds to the value of the t -th element of the vector \mathbf{z} , the initial observation is $a = 0, 1$, while $s = 1, \dots, t$ and $t = 1, \dots, T_i$ are the total score and the length of the vector configuration \mathbf{y}_i , respectively. This general structure can be easily adapted to the three QE models defining different functions $g_{t,a,s}(\boldsymbol{\phi}, \psi)$ that may be recursively computed. The main difference with respect to the static case is that, further to the total score and the time dimension, we take into account not only the initial observation but also the t -th element of vector \mathbf{z} .

3.2.1 Simplified QE Model

Consider the simplified version of the QE model described by (12) and the corresponding normalizing constant expressed as the denominator of (11). In general, we define a single element of the sum at the rhs of Equation (18) as

$$g_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \sum_{\mathbf{z}: z_t = s, z_t = v} \exp \left[\sum_{u=1}^t z_u \phi_u + \left(a z_1 + \sum_{u=2}^t z_{u-1} z_u \right) \psi \right],$$

where $\phi_u = \mathbf{x}'_{iu} \boldsymbol{\eta}$. Following the formulation above, the recursion to compute $g_{t,a,s,v}(\boldsymbol{\phi}, \psi)$ is based on the following steps:

1. for $t = 1$, let

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

with $a = 0, 1$;

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

$$g_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} 1, & s = 0, v = 0, \\ g_{t-1,a,s,0}(\boldsymbol{\phi}, \psi) + g_{t-1,a,s,1}(\boldsymbol{\phi}, \psi), & s = 1, \dots, t-1, v = 0, \\ g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi) \exp(\phi_t) \\ \quad + g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi) \exp(\phi_t + \psi), & s = 1, \dots, t, v = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

3.2.2 Modified QE Model

Following the same strategy as above, the recursion for the modified QE model can be adopted by defining

$$g_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \sum_{\mathbf{z}: z_+ = s, z_t = v} \exp \left[\sum_{u=1}^t z_u \phi_u + \left(\mathbb{1}\{z_1 = a\} + \sum_{u=2}^t \mathbb{1}\{z_u = z_{u-1}\} \right) \psi \right].$$

The recursion for computing $g_{t,a,s,v}(\boldsymbol{\phi}, \psi)$ is as follows:

- for $t = 1$, let

$$g_{1,a,s,v}(\boldsymbol{\phi}, \psi) = \{ \exp[(\mathbb{1} - a)\psi], s = v = 0, \exp(\phi_1 + a\psi), s = v = 1, 0, \text{ otherwise,}.$$

with $a = 0, 1$;

- for $t = 2, \dots, T_i$, compute

$$g_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \begin{cases} g_{t-1,a,s,0}(\boldsymbol{\phi}, \psi) \exp(\psi) + g_{t-1,a,s,1}(\boldsymbol{\phi}, \psi), & s = 0, \dots, t-1, v = 0, \\ g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \psi) \exp(\phi_t) + g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \psi) \exp(\phi_t + \psi), & s = 1, \dots, t, v = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

3.2.3 Approximating QE Model

Finally, for the approximating QE model leading to Eq. (14), we define

$$g_{t,a,s,v}(\boldsymbol{\phi}, \psi) = \sum_{\mathbf{z}: z_+ = s, z_t = v} \left[\sum_{u=1}^t z_u \phi_u + \sum_{u=1}^t v_u z_{u-1} + \left(\sum_{u=1}^t z_{u-1} z_u \right) \gamma \right],$$

where $z_0 = a$, $\phi_u = \mathbf{x}'_{iu} \boldsymbol{\beta}$, and $v_u = -\bar{r}_{iu} \gamma$.

Exploiting the same structure of Eq. (18), the recursive computation of $g_{t,a,s,v}(\boldsymbol{\phi}, \gamma)$ can be performed as follows:

- for $t = 1$, let

$$g_{1,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} \exp(v_1 a), & s = v = 0 \\ \exp[\phi_1 + a(v_1 + \gamma)], & s = v = 1, \\ 0, & \text{otherwise,} \end{cases}$$

with $a = 0, 1$;

- for $t = 2, \dots, T_i$, compute

$$g_{t,a,s,v}(\boldsymbol{\phi}, \gamma) = \begin{cases} 1, & s = 0, v = 0, \\ g_{t-1,a,s,0}(\boldsymbol{\phi}, \gamma) + g_{t-1,a,s,1}(\boldsymbol{\phi}, \gamma) \exp(v_t), & s = 1, \dots, t-1, v = 0, \\ g_{t-1,a,s-1,0}(\boldsymbol{\phi}, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}(\boldsymbol{\phi}, \gamma) \exp(\phi_t + \gamma + v_t), & s = 1, \dots, t, v = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

4 Computational Complexity

We now focus on the comparison between the different methods to evaluate the conditional probability of a response configuration in terms of computational complexity. For what concerns the standard method based on Equation (11) applied for each unit $i = 1, \dots, n$, this is based on a sum extended to all the vectors \mathbf{z} , whose number depends on the time dimension and on the total score, as given by the binomial coefficient $k = \binom{T_i}{y_{i+}}$. When T_i is large (e.g., $T \geq 15$), the number of operations k may become huge as well as the number of operations required for the computation of the normalizing constant. Figure 1 shows the total number of vectors \mathbf{z} for a small grid of values of T_i and y_{i+} . Obviously, when the number k of such vectors becomes too large, the computation of (11) is no longer feasible.

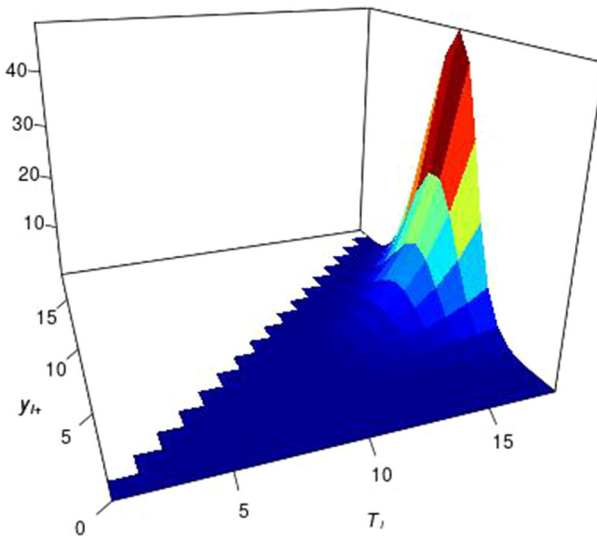


Fig. 1 Numbers of vectors \mathbf{z} (in thousands) as a function of T_i and y_{i+}

The proposed recursion overcomes the computational problem described above. Again, the operations required depend on T_i and y_{i+} , but their order is reduced because basic sums and products are computed, at most, for each value of the total score s in $\{1, \dots, t\}$, and repeated for t times, with t varying in $\{1, \dots, T_i\}$. This aspect results in a significant difference in terms of computational time.

For a clearer comparison between different methods, the computational time of the proposed recursive algorithm is empirically evaluated by a simple Monte Carlo simulation that is described in this section. In this regard consider that experiments are run on a computer with n. 2 Intel Xeon CPU E5-2640 v4 2.40GHz, 250GB of RAM, with Debian GNU/Linux “bullseye”/sid as operating system. Results are obtained using R 4.0.3 and the custom code included in the `cquad` package.

The simulation design is similar to the one adopted by Honoré & Kyriazidou (2000). In particular, data are generated from the DL model according to expression (8), where x_{it} is an exogenous regressor following a Gaussian distribution with zero mean and variance $\pi^2/3$, the parameter β is equal to 1, and the state dependence parameter γ is equal to 0.5. Individual effects are generated as $\alpha_i = \frac{1}{4} \sum_{t=0}^3 x_{it}$ and the Monte Carlo replications are 50. Furthermore, the initial observation has distribution

$$p(y_{i0}|\alpha_i, x_{i0}) = \frac{\exp[y_{i0}(\alpha_i + x_{i0}\beta)]}{1 + \exp(\alpha_i + x_{i0}\beta)}.$$

Finally, we consider $n = 250$ individuals observed for a different number of time occasions. In order to simplify the interpretation of the results, the study is based on

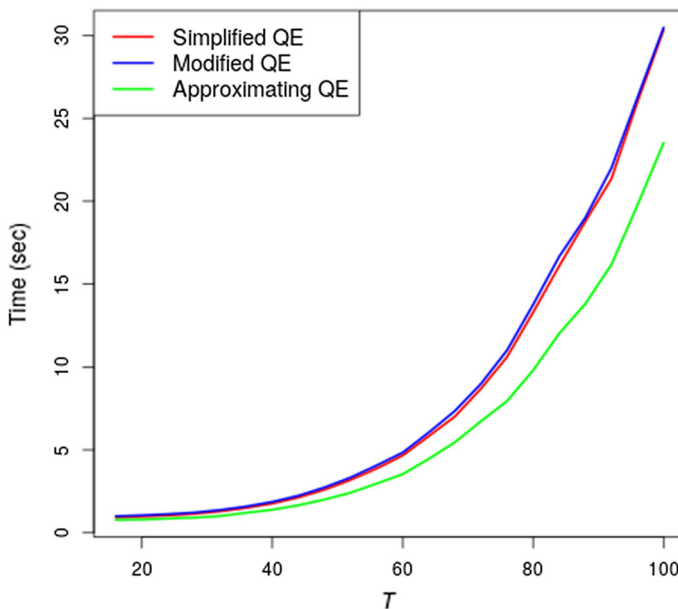


Fig. 2 CPU time of the recursion: QE models

simulated datasets of 249 subjects observed for $T_i = 5$ time occasions and only one subject, the last, observed for $T > 5$ time occasions, where T is allowed to vary in different ranges.

Within the simulation study we measure how the CPU time required by the Newton-Raphson algorithm for CML estimation varies according to the time dimension of the panel. In this case we evaluate T in a grid between 16 and 100 with steps of 4. For each value of T , the average CPU time taken by the maximization over the replications are shown in Fig. 2. The CPU time obviously increases in T and it is similar for the three models considered.

With a second experiment, we aim at comparing the required time for the computation of the log-likelihood function with both the standard algebraic operation and using the recursive algorithm.

In this case, we make use of a single draw from the simulation design described above, with $T_i = 5$ for $i = 1, \dots, 249$. The outcome for the 250-th subject is arbitrarily given by a sequence that alternates 0 and 1, so as to rule out the randomness given by the data generating process and, at the same time, to maximize the number of vectors, namely k , controlling for the total score. Further, the initial observation is set to 1 and T ranges in $\{8, \dots, 18\}$.

Figures 3, 4, and 5 represent the computational time required for the log-likelihood maximization of the simplified, modified, and approximating QE models, respectively. In terms of computational load, the results for the modified QE model are entirely comparable with those for the simplified QE model, while the

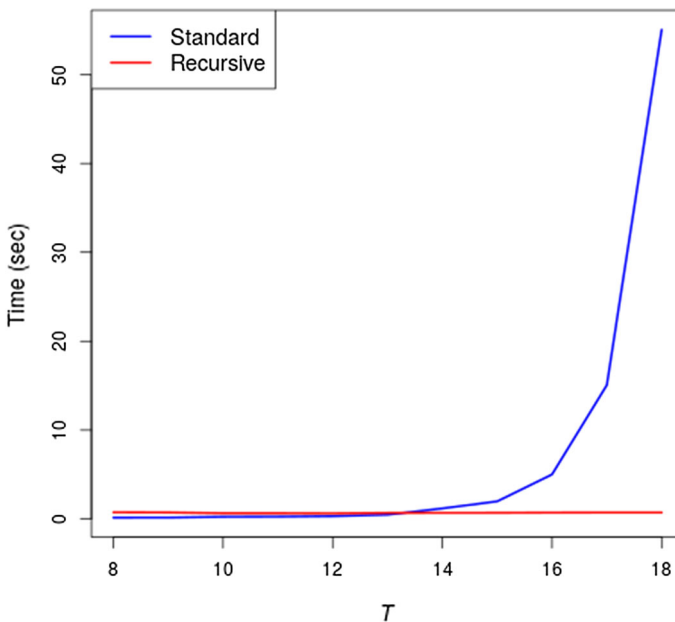


Fig. 3 CPU time comparison of algebraic and recursive computation: Simplified QE

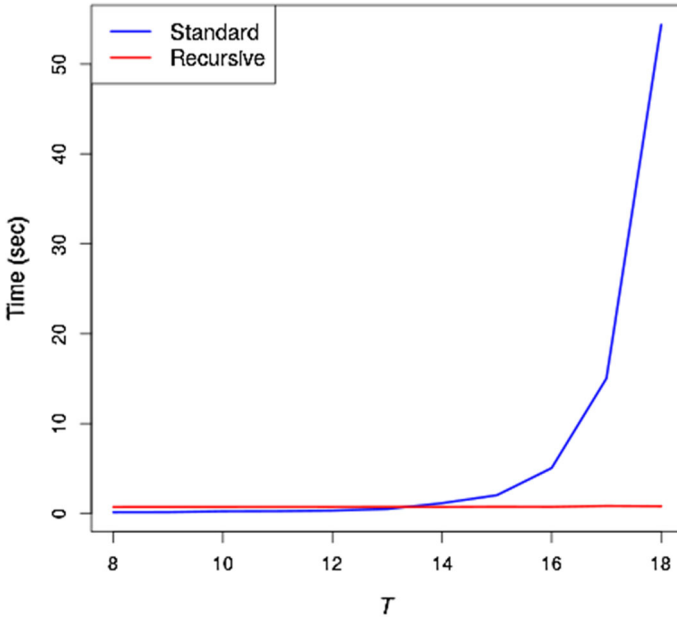


Fig. 4 CPU time comparison of algebraic and recursive computation: Modified QE

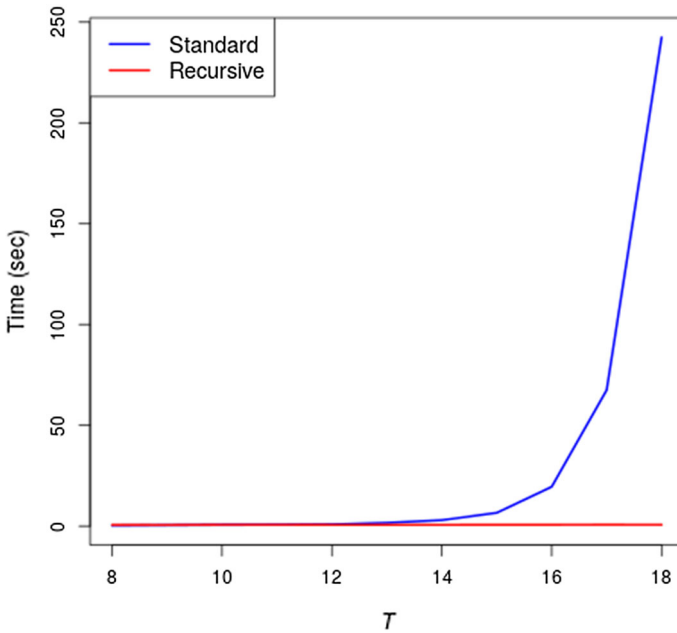


Fig. 5 CPU time comparison of algebraic and recursive computation: Approximating QE

approximating QE is more demanding since it consists in a two-step CML estimator.

Each figure reports the time required by the algebraic operations (“Standard”) and the recursion (“Recursive”), in which we can see that the former exhibits an exponential pace as T approaches 18 while the latter remains stable. The time taken by the algebraic computation of the “one-step” QE models grows up to 1 minute while the PCML routine takes more than 4 minutes for $T = 18$. For what concerns the algorithm, the computational time is always lower than one second for all the time dimensions considered.

5 Application: Brand Loyalty

The proposed methodology is applied to real data on consumers loyalty to two different dominant yogurt brands. The analysis here performed replicates the example provided by Chintagunta et al. (2001). Differently from the original work, though, the present application is based on a sample dataset provided by A.C. Nielsen² that contains information about yogurt purchases made by individuals observed for a period of about two years. This dataset has already been used by Jain et al. (1994).

Purchases concern four brands: Yoplait, Dannon, Nordica, and Weight. As in the original work, we only keep purchases of the two brands with the largest market share in this dataset, namely Dannon and Yoplait. The sample consists of an unbalanced panel of 100 consumers, for a total of 1,788 observations. Time occasions differ among individuals and range between a minimum of 1 and a maximum of 161.

The response variable for the model considered is set to 1 when a consumer chose Dannon. We also include two exogenous explanatory variables. The first one is the log-difference between the prices of the two products (Price). The second one (Featured) is a categorical variable built as the difference between two dummy variables, one per brand, recording whether the brand is advertised in newspapers. We expect that a small relative price and advertisement should have a positive impact on the demand of the goods and then on the probability of a purchase for a brand. Moreover, stickiness in consumer habits require to be taken into account by means of a dynamic specification of the model, that is, by including the lagged dependent variable in the set of covariates. Chintagunta et al. (2001) discuss different model specifications but only the DL model allows us to properly include individual fixed effects and account for true state dependence.

The models here considered are extremely useful to represent consumers behavior and their loyalty to a brand (Chintagunta et al., 2001). First of all, from a statistical perspective, it is straightforward to represent brand choices as a binary variable being 1 or 0 according to whether a product is chosen or not. Secondly, the panel structure of the data allows us to take into account some individual time-constant unobserved characteristics. Since consumers are observed for a relatively

² Data are publicly available in the R package `Ecdat` (<https://cran.r-project.org/web/packages/Ecdat/index.html>).

short span of time, in which their purchases are recorded, from a theoretical perspective it is reasonable to consider some individual unobservable characteristics as time-invariant.

In the original example, Chintagunta et al. (2001) specify the DL model and consider different estimation approaches: the ML estimator of the pooled model (with no individual intercepts), the CML estimator where the lagged dependent variable is treated as exogenous, the weighted CML (WCML) estimator by Honoré & Kyriazidou (2000), and different random-effects specifications. The proposed algorithm allows us to extend the analysis to QE models as well. In particular, here we formulate the brand choice model as a simplified QE model and as the QE approximation of the DL model. The parameters of these model formulations are estimated by CML and PCML, respectively. Estimation results are reported in Table 1, together with the pooled logit model, the CML, and the WCML.

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

Finally, it is worth stressing that there is no way to compare the computational time required by the algorithm for the QE models and their standard matrix

Table 1 Estimation results: DL and QE models for brand choice

	DL Pooled ML	DL* CML	DL WCML**	Sim. QE CML	App. QE PCML
Dannon _{<i>t</i>-1}	4.634 (0.183)	1.715 (0.317)	1.778 (0.208)	2.118 (0.221)	2.326 (0.389)
Price	-3.373 (0.384)	-3.565 (0.771)	-3.264 (0.654)	-3.264 (0.514)	-3.390 (0.702)
Featured	0.362 (0.275)	0.739 (0.490)	-0.121 (0.456)	0.440 (0.317)	0.723 (0.438)
Time (sec)	<1	1.399	2.384	31.996	35.335

* Lagged dependent variable treated as exogenous.

** Bandwidth parameter for WCML, $c = 10$

computation since, in the latter case, the estimation is not feasible due to the large numbers of time occasions.

6 Conclusions

This work provides a novel way to compute the conditional likelihood functions of the QE model and its extensions. By means of a recursive algorithm, the standard computational burden of algebraic calculations is relieved for applications that involve a panel dataset with large time dimensions.

In particular, the application of the algorithm here proposed is relevant in empirical works making use of unbalanced panel datasets. The presence of subjects observed for a few time occasions makes the conditional inference approach more appealing than the ML estimator, which is inconsistent with fixed T . At the same time, the proposed algorithm dramatically limits the burden of computing the conditional log-likelihood, which can be sizable even if only few subjects are observed for a long time span.

A Monte Carlo experiment shows how the proposed algorithm outperforms standard computation and it should enlarge the applicability of the considered models. In this regard, an application to real data concerning brand loyalty has been proposed with the additional interest of showing a set of results that would have been otherwise impossible, providing practitioners with a wider range of estimation tools for empirical analysis.

Finally, the conditional inference approach can be easily extended to ordered data, as a fixed-effects ordered logit model can be estimated by relying on a set of fixed-effects binary logit models, which are referred to binary variables with categories obtained from a dichotomization of the original responses (Chamberlain, 1980; Baetschmann et al., 2015).

Appendix: Derivatives

In the following, we report the first and second derivatives for the relevant functions of the proposed recursive algorithm for the static logit, simplified, modified, and approximating QE models. These derivatives are then used to compute the score vector and Information matrix involved in the steps of the Newton-Raphson algorithm, in order to maximize the conditional log-likelihoods.

Static Logit Model

Concerning computation of the derivatives presented in Sect. 3.1 we can exploit the same recursive structure presented in Eq. (17), so that:

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

$$f_{1,0}^{(h)}(\boldsymbol{\phi}) = 0, \quad h = 1, \dots, T_i,$$

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

and the second derivatives are

$$f_{1,0}^{(h,j)}(\boldsymbol{\phi}) = 0, \quad h, j = 1, \dots, T_i,$$

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

2. for $t = 2, \dots, T_i$ and $s = 1, \dots, t$, are the first derivatives are

$$f_{t,0}^{(h)}(\boldsymbol{\phi}) = 0, \quad h = 1, \dots, T_i,$$

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

and second derivatives are

$$f_{t,0}^{(h,j)}(\boldsymbol{\phi}) = 0, \quad h, j = 1, \dots, T_i,$$

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

Simplified QE Model

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

$$g_{t,s}^{(h)}(\boldsymbol{\phi}) = \frac{\partial g_{t,s}(\boldsymbol{\phi})}{\partial \phi_h},$$

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

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

Regarding the first derivatives of the function defined in (19), we exploit the same recursion:

1. when $t = 1$ compute

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

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

2. when $t = 2, \dots, T_i$ consider the following cases:

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

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

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

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

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

Regarding the second derivatives, we have:

1. when $t = 1$ compute

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

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

2. when $t = 2, \dots, T_i$ consider the following cases:

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

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

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

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

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

Modified QE Model

Regarding the first derivatives of the function reported in Eq. (20), we exploit the same recursion:

1. with $t = 1$

- for $s = v = 0$ compute

$$g_{1,a,s,v}^{(h)}(\phi, \psi) = \begin{cases} 0, & h = 1, \\ (1-a) \exp[\phi_1 + (1-a)\psi], & h = T_i + 1; \end{cases}$$

- for $s = v = 1$ compute

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

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

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

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

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

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

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

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

Regarding the second derivatives, we have:

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

$$g_{1,a,s,v}^{(h,j)}(\phi, \psi) = (1 - a) \exp[(1 - a)\psi], \quad h, j = T_i + 1;$$

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

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

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

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

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

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

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

$$\begin{aligned}
 &g_{t,a,s,v}^{(h,j)}(\phi, \psi) \\
 = &\left\{ \begin{array}{ll}
 g_{t-1,a,s-1,0}^{(h,j)}(\phi, \psi) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \psi) \exp(\phi_t + \psi), & h, j = 1, \dots, t-1, \\
 g_{t,a,s,v}^{(h)}(\phi, \psi), & h = 1, \dots, t-1, j = t, \\
 g_{t-1,a,s-1,0}^{(h,j)}(\phi, \psi) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \psi) \exp(\phi_t + \psi) \\
 + g_{t-1,a,s-1,1}^{(h)}(\phi, \psi) \exp(\phi_t + \psi), & h = 1, \dots, t-1, j = T_i + 1, \\
 g_{t,a,s,v}^{(j)}(\phi, \psi), & h = t, j = 1, \dots, t, T_i + 1, \\
 g_{t-1,a,s-1,0}^{(h,j)}(\phi, \psi) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \psi) \exp(\phi_t + \psi) \\
 + g_{t-1,a,s-1,1}^{(j)}(\phi, \psi) \exp(\phi_t + \psi), & h = T_i + 1, j = 1, \dots, t-1 \\
 g_{t,a,s,v}^{(h)}(\phi, \psi), & h = T_i + 1, j = t, \\
 g_{t-1,a,s-1,0}^{(h,j)}(\phi, \psi) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \psi) \exp(\phi_t + \psi) \\
 + g_{t-1,a,s-1,1}^{(h)}(\phi, \psi) \exp(\phi_t + \psi) + g_{t-1,a,s-1,1}^{(j)}(\phi, \psi) \exp(\phi_t + \psi) \\
 + g_{t-1,a,s-1,1}^{(h)}(\phi, \psi) \exp(\phi_t + \psi), & h = T_i + 1, j = T_i + 1;
 \end{array} \right.
 \end{aligned}$$

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

Approximating QE model

Regarding the first derivatives of the function in Eq. (21), compute:

1. with $t = 1$,

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

$$g_{1,a,s,v}^{(h)}(\phi, \gamma) = \begin{cases} a \exp(av_1)(-\bar{r}_{i1}), & h = T_i + 1, \\ 0, & \text{otherwise;} \end{cases}$$

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

$$g_{1,a,s,v}^{(h)}(\phi, \gamma) = \begin{cases} \exp(\phi_1 + a(v_1 + \gamma)), & h = 1, \\ a \exp(\phi_1 + a(v_1 + \gamma))(1 - \bar{r}_{i1}), & h = T_i + 1; \end{cases}$$

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

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

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

$$g_{t,a,s,v}^{(h)}(\phi, \gamma) = \begin{cases} g_{t-1,a,s,0}^{(h)}(\phi, \gamma) + g_{t-1,a,s,1}^{(h)}(\phi, \gamma) \exp(v_t), & h = 1, \dots, t, \\ g_{t-1,a,s,0}^{(h)}(\phi, \gamma) + g_{t-1,a,s,1}^{(h)}(\phi, \gamma) \exp(v_t) + g_{t-1,a,s,1}^{(h)}(\phi, \gamma) \exp(v_t)(-\bar{r}_{it}), & h = T_i + 1; \end{cases}$$

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

$$g_{t,a,s,v}^{(h)}(\phi, \gamma) = \begin{cases} g_{t-1,a,s-1,0}^{(h)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) + g_{t-1,a,s-1,0}^{(h)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma), & h = 1, \dots, t \\ g_{t-1,a,s-1,0}^{(h)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) + g_{t-1,a,s-1,1}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it}), & h = T_i + 1; \end{cases}$$

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

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

1. for $t = 1,$

- with $v = 0$ compute

$$g_{1,a,s,0}^{(h,j)}(\phi, \psi) = a \exp(av_1)(-\bar{r}_{i1})^2, \quad h, j = T_i + 1;$$

- while for $v = 1$ we have

$$g_{1,a,s,1}^{(h,j)}(\phi, \psi) = \begin{cases} \exp(\phi_1 + a(v_1 + \gamma)), & h = j = 1, \\ a \exp(\phi_1 + a(v_1 + \gamma))(1 - \bar{r}_{i1}), & h = 1, j = T_i + 1, \\ a \exp(\phi_1 + a(v_1 + \gamma))(1 - \bar{r}_{i1}), & h = T_i + 1, j = 1, \\ a \exp(\phi_1 + a(v_1 + \gamma))(1 - \bar{r}_{i1})^2, & h, j = T_i + 1; \end{cases}$$

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

2. with $t = 2, \dots, T_i$:

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

$$g_{t,a,s,v}^{(h,j)}(\phi, \gamma) = \begin{cases} g_{t-1,a,s,0}^{(h,j)}(\phi, \gamma) + g_{t-1,a,s,1}^{(h,j)}(\phi, \gamma) \exp(v_t), & h, j = 1, \dots, t \\ g_{t-1,a,s,0}^{(h,j)}(\phi, \gamma) + g_{t-1,a,s,1}^{(h,j)}(\phi, \gamma) \exp(v_t) \\ \quad + g_{t-1,a,s,1}^{(j)}(\phi, \gamma) \exp(v_t)(-\bar{r}_{it}), & h = T_i + 1, j = 1, \dots, t \\ g_{t-1,a,s,0}^{(h,j)}(\phi, \gamma) + g_{t-1,a,s,1}^{(h,j)}(\phi, \gamma) \exp(v_t) \\ \quad + g_{t-1,a,s,1}^{(h)}(\phi, \gamma) \exp(v_t)(-\bar{r}_{it}), & h = 1, \dots, t, j = T_i + 1, \\ g_{t-1,a,s,0}^{(h,j)}(\phi, \gamma) + g_{t-1,a,s,1}^{(h,j)}(\phi, \gamma) \exp(v_t) \\ \quad + g_{t-1,a,s,1}^{(h)}(\phi, \gamma) \exp(v_t)(-\bar{r}_{it}) + g_{t-1,a,s,1}^{(j)}(\phi, \gamma) \exp(v_t)(-\bar{r}_{it}) \\ \quad \quad + g_{t-1,a,s,1} \exp(v_t)(-\bar{r}_{it})^2, & h = T_i + 1, j = T_i + 1; \end{cases}$$

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

$$g_{t,a,s,v}^{(h,j)}(\phi, \gamma) = \begin{cases} g_{t-1,a,s-1,0}^{(h,j)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad + g_{t-1,a,s-1,0}^{(h)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad + g_{t-1,a,s-1,0}^{(j)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad \quad + g_{t-1,a,s,0}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s,1}(\phi, \gamma) \exp(\phi_t + v_t + \gamma), & h, j = 1, \dots, t, \\ g_{t-1,a,s-1,0}^{(h,j)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad + g_{t-1,a,s-1,1}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it}) + g_{t-1,a,s-1,0}^{(j)}(\phi, \gamma) \exp(\phi_t) \\ \quad \quad + g_{t-1,a,s-1,1}^{(j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad \quad + g_{t-1,a,s,1}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it}), & h = 1, \dots, t, j = T_i + 1, \\ g_{t-1,a,s-1,0}^{(h,j)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad + g_{t-1,a,s-1,1}^{(j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it}) \\ \quad + g_{t-1,a,s-1,0}^{(h)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad \quad + g_{t-1,a,s,1}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it}), & h = 1, \dots, t, j = T_i + 1, \\ g_{t-1,a,s-1,0}^{(h,j)}(\phi, \gamma) \exp(\phi_t) + g_{t-1,a,s-1,1}^{(h,j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma) \\ \quad + g_{t-1,a,s-1,0}^{(h)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it}) \\ \quad + g_{t-1,a,s-1,1}^{(j)}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it}) \\ \quad \quad + g_{t-1,a,s-1,1}(\phi, \gamma) \exp(\phi_t + v_t + \gamma)(1 - \bar{r}_{it})^2, & h = T_i + 1, j = T_i + 1; \end{cases}$$

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

Funding F. Bartolucci acknowledges the financial support from the grant “Partial effects in econometric models for binary longitudinal data based on quadratic exponential distributions” of the University of Perugia (RICBASE2018).

Data Availability Data used in Sect. 5 are publicly available at <https://cran.r-project.org/package=Ecdata>.

Code Availability Custom code included in the R package <https://cran.r-project.org/package=cquad>.

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Agresti, A. (2003). *Categorical Data Analysis*, (Vol. 482). Wiley.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B*, 32, 283–301.
- Andersen, E. B. (1971). The asymptotic distribution of conditional likelihood ratio tests. *Journal of the American Statistical Association*, 66, 630–633.
- Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society: Series B*, 34, 42–54.
- Arellano, M., & Hahn, J. (2016). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. *Global Economic Review*, 45, 251–274.
- Arellano, M., & Honoré, B. (2001). Panel data models: some recent developments. In: *Handbook of Econometrics*, vol 5, Elsevier, pp 3229–3296.
- Baetschmann, G., Staub, K. E., & Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A*, 178, 685–703.
- Bartolucci, F., & Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a \sqrt{n} -consistent conditional estimator. *Econometrica*, 78, 719–733.
- Bartolucci, F., & Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data. *Journal of Econometrics*, 170, 102–116.
- Bartolucci, F., & Pennoni, F. (2007). On the approximation of the quadratic exponential distribution in a latent variable context. *Biometrika*, 94, 745–754.
- Bartolucci, F., & Pignini, C. (2017). cquad: An R and Stata package for conditional maximum likelihood estimation of dynamic binary panel data models. *Journal of Statistical Software*, 78(7)
- Bartolucci, F., Bacci, S., & Gnaldi, M. (2015). *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata*. Chapman and Hall/CRC.
- Bartolucci, F., Bellio, R., Salvan, A., & Sartori, N. (2016). Modified profile likelihood for fixed-effects panel data models. *Econometric Reviews*, 35, 1271–1289.
- Bartolucci, F., Nigro, V., & Pignini, C. (2018). Testing for state dependence in binary panel data with individual covariates by a modified quadratic exponential model. *Econometric Reviews*, 37, 61–88.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36, 192–225.
- Bester, C. A., & Hansen, C. (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics*, 27, 131–148.
- Carro, J. M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, 140, 503–528.

- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, *47*, 225–238.
- Chamberlain, G. (1993). *Feedback in panel data models*. Harvard-Institute of Economic Research: Tech. rep.
- Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal*, *20*, S1–S13.
- Chintagunta, P., Kyriazidou, E., & Perktold, J. (2001). Panel data analysis of household brand choices. *Journal of Econometrics*, *103*, 111–153.
- Cox, D. R. (1972a). The analysis of multivariate binary data. *Applied Statistics*, *21*, 113–120.
- Cox, D. R. (1972b). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, *34*, 187–202.
- De Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, *12*, 775–799.
- Dhaene, G., & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, *82*, 991–1030.
- Diggle, P. J., Heagerty, P., Liang, K. Y., Heagerty, P. J., Zeger, S., et al. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
- Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, *150*, 71–85.
- Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, *85*, 1033–1063.
- Gustafsson, J. E. (1980). A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement*, *40*, 377–385.
- Hahn, J., & 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.
- Hambleton, R. K., & Swaminathan, H. (1986). *Item Response Theory: Principles and Applications*. Springer Science & Business Media.
- Heckman, J. J. (1981). Heterogeneity and state dependence. In: *Studies in labor markets*, University of Chicago Press, pp 91–140.
- Honoré, B. E., & Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, *68*, 839–874.
- Howard, S. (1972). Remark on the paper by Cox, D. R. (1972): Regression methods and life tables. *Journal of the Royal Statistical Society: Series B*, *34*, 187–202.
- Hsiao, C. (2014). *Analysis of Panel Data*. Cambridge University Press.
- Jain, D. C., Vilcassim, N. J., & Chintagunta, P. K. (1994). A random-coefficients logit brand-choice model applied to panel data. *Journal of Business & Economic Statistics*, *12*, 317–328.
- Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics*, *36*(4), 705–713.
- Krailo, M., & Pike, M. (1984). Algorithm AS 196: conditional multivariate logistic analysis of stratified case-control studies. *Journal of the Royal Statistical Society Series C*, *33*, 95–103.
- Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*, 391–413.
- Levin, B. (1987). Conditional likelihood analysis in stratum-matched retrospective studies with polytomous disease states. *Communications in Statistics: Simulation and Computation*, *16*, 699–718.
- Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*(413), 96–107.
- Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, *16*, 1–32.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research.

- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, 4, 321–333.
- Smith, P., Pike, M., Hill, A., Breslow, N., & Day, N. (1981). Algorithm AS 162: Multivariate conditional logistic analysis of stratum-matched case-control studies. *Journal of the Royal Statistical Society: Series C*, 30, 190–197.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, MIT Press Books, (Vol. 1). The MIT Press.
- Zhang, D., & Davidian, M. (2004). Likelihood and conditional likelihood inference for generalized additive mixed models for clustered data. *Journal of Multivariate Analysis*, 91(1), 90–106.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.