Partial effects estimation for fixed-effects logit panel data models

Bartolucci, Francesco and Pigini, Claudia

18 February 2019

Partial effects estimation for fixed-effects logit 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

Abstract

We propose a multiple step procedure to estimate Average Partial Effects (APE) in fixed-effects panel logit models. Because the incidental parameters problem plagues the APEs via both the inconsistent estimates of the slope and individual parameters, we reduce the bias by evaluating the APEs at a fixed-T consistent estimator for the slope coefficients and at a bias corrected estimator for the unobserved heterogeneity. The proposed estimator has bias of order $O(T^{-2})$ as $n \to \infty$ and performs well in finite sample, even when n is much larger than T. We provide a real data application based on the labor supply of married women.

1 Introduction

Practitioners who estimate binary choice models are often interested in quantifying the effect of some regressor x on the response probability, other things being equal. Moreover, with the availability of panel data, the fixed-effects approach allows for the estimation of partial effects of covariates that may be correlated with the individual specific unobserved heterogeneity in a nonparametric manner.

The Maximum Likelihood (ML) estimator of fixed-effects binary choice models, however, is consistent only as $T \to \infty$ and otherwise suffers form the well-known incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000).[1] With fixed T, the plug-in estimator of the Average Partial Effect (APE) of some covariate on the response probability is also plagued by the incidental parameters problem, which gives rise to two sources of bias: one is introduced by the ML estimator of the individual effects, consistent only as $T \to \infty$; the other is carried over by the ML estimates of the slope parameters, that are affected by the bias in the estimated subject-specific intercepts as they are not informationally orthogonal.

We propose a multiple step procedure to estimate the APE in fixed-effects panel logit models with a reduced order of bias. The bias introduced by the estimated slopes is removed by using the fixed-T consistent Conditional Maximum Likelihood (CML) estimator for parameters of the fixed-effects logit model, for which the incidental parameters problem can be solved by conditioning on simple sufficient statistics for the individual intercepts (Andersen, 1970; Chamberlain, 1980). The bias that comes with the ML estimates of the individual intercepts is reduced from $O(T^{-1})$ to $O(T^{-2})$ by defining the ML estimator for the individual unobserved heterogeneity as the solution to the modified score function put forward by Firth (1993).

The proposed procedure, however, cannot be extended directly to the dynamic logit (Hsiao, 2005), for which CML inference for the slope parameters is not viable in a simple form. This is overcome by Bartolucci and Nigro (2010), who propose a Quadratic Exponential (QE) formulation (Cox, 1972) to model dynamic binary panel data, that has the advantage of admitting sufficient statistics for the individual intercepts. Furthermore, Bartolucci and Nigro (2012) propose a QE model, that approximates more closely the dynamic logit model, the parameters of which can easily be estimated by Pseudo CML (PCML). We therefore extend the proposed procedure to include PCML estimates in the APEs when a dynamic logit is specified.

Several contributions deal with bias reduction techniques for the ML estimators of fixed-effects binary choice models. Some of them provide bias corrections for the APEs as well, along the same lines of the corrections proposed for the slopes. Analytical corrections

---

[1]We focus on large n and large T perspective, as APEs are often not point identified with fixed T (Chernozhukov et al., 2013).

are provided by Fernández-Val (2009), whose derivations are based on general results for static (Hahn and Newey, 2004) and dynamic (Hahn and Kuersteiner, 2011) nonlinear panel data models. An alternative bias correction method for the APE estimator relies on the panel jackknife. A general procedure for nonlinear static panel data models is proposed by Hahn and Newey (2004), whereas a split-panel jackknife estimator is developed by Dhaene and Jochmans (2015) for dynamic models.

As it happens with the proposed method, the analytical and jackknife corrections reduce the order of the bias for the APE from $O(T^{-1})$ to $O(T^{-2})$. However, these APE estimators depend on some bias corrected estimator for the slope coefficients which can be shown to have correct confidence intervals if T grows faster than $n^{1/3}$ (Hahn and Newey, 2004). This is not required in our case. In fact, we show by simulation that the proposed APE estimator, although large-T consistent, performs well in finite samples even when n is much larger than T.

The rest of the paper is organized as follows: in Section 2 we briefly discuss the incidental parameters problem and how it affects the APEs estimator; in Section 3 we illustrate the proposed methodology, its extension to accommodate the dynamic logit model, and briefly recall the alternative bias correction strategies; in Section 4 we investigate the finite sample performance of the proposed estimator and compare it with that of the panel jackknife; in Section 5 we provide a real data application based on labor supply of married women. Finally, Section 6 concludes.

2 Average partial effects and the incidental parameters problem

We consider n units, indexed with $i = 1, \ldots, n$, observed at time occasions $t = 1, \ldots, T$. Let y_{it} be the binary response variable for unit i at occasion t and $\boldsymbol{x}_{it}$ the corresponding vector of K covariates. We assume that both y_{it} and $\boldsymbol{x}_{it}$ are independent across i and T. Consider the logit formulation

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

where α_i is the individual specific intercept, $\boldsymbol{x}_{it}$ is vector of strictly exogenous covariates, and $\boldsymbol{\beta}$ collects the regression parameters.

The fixed-effects estimator is obtained by Maximum Likelihood (ML), treating each individual effect α_i as a parameter to be estimated. The ML estimator of $\boldsymbol{\beta}_0$ is obtained

by concentrating out the α_i as the solution to

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \ln p(y_{it}|\boldsymbol{x}_{it}; \hat{\alpha}_i(\boldsymbol{\beta}), \boldsymbol{\beta}),$$

$$\hat{\alpha}_i(\boldsymbol{\beta}) = \underset{\alpha_i}{\arg\max} \sum_{t=1}^{T} \ln p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}).$$

Notice that here $\hat{\alpha}_i(\boldsymbol{\beta})$ depends on the data only through $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})$.

Because the estimation noise in $\hat{\alpha}_i(\boldsymbol{\beta})$ disappears only as $T \to \infty$, the ML estimator of $\hat{\boldsymbol{\beta}}$ is not consistent for $\boldsymbol{\beta}_0$ with T fixed and only $n \to \infty$, that is $\underset{n\to\infty}{\text{plim}} \hat{\boldsymbol{\beta}} \equiv \boldsymbol{\beta}_T \neq \boldsymbol{\beta}_0$. This is the well-known incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000). To clarify this, consider any function $m(\boldsymbol{y}_i, \boldsymbol{X}_i, \alpha_i)$ and let $\mathrm{E}_n[m(\boldsymbol{y}_i, \boldsymbol{X}_i, \alpha_i)] \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} m(\boldsymbol{y}_i, \boldsymbol{X}_i, \alpha_i)$, where α_i is treated as fixed. From standard extremum estimator properties, it follows that, with T fixed and as $n \to \infty$, $\boldsymbol{\beta}_T$ is be obtained as

$$\boldsymbol{\beta}_T = \underset{\boldsymbol{\beta}}{\arg\max} \, \mathrm{E}_n \left[\sum_{t=1}^{T} \ln p(y_{it}|\boldsymbol{x}_{it}; \hat{\alpha}_i(\boldsymbol{\beta}), \boldsymbol{\beta}) \right],$$

whereas $\boldsymbol{\beta}_0$ follows from

$$\boldsymbol{\beta}_0 = \underset{\boldsymbol{\beta}}{\arg\max} \, \mathrm{E}_n \left[\sum_{t=1}^{T} \ln p(y_{it}|\boldsymbol{x}_{it}; \alpha_i(\boldsymbol{\beta}), \boldsymbol{\beta}) \right],$$

where $\alpha_i(\boldsymbol{\beta})$ maximizes $\mathrm{E}_T[\ln p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta})]$. From the expressions above it is clear that the problem arises from $\hat{\alpha}_i(\boldsymbol{\beta}) \neq \alpha_i(\boldsymbol{\beta})$ with fixed T. Moreover, Hahn and Newey (2004) show that $\boldsymbol{\beta}_T = \boldsymbol{\beta}_0 + B/T + O(T^{-2})$. If, instead, $T \to \infty$, then $\hat{\alpha}_i(\boldsymbol{\beta}) \to \alpha_{i0}$, with $\alpha_{i0} = \alpha_i(\boldsymbol{\beta})$, and $\boldsymbol{\beta}_T \to \boldsymbol{\beta}_0$. If both $n, T \to \infty$, $\hat{\boldsymbol{\beta}}$ will be consistent and asymptotically normal. However, Hahn and Newey (2004) show that the asymptotic distribution of ML estimator will not be centered at its probability limit if n grows faster than T.

The incidental parameters problem severely affects the estimation of APEs as well, that are usually of interest to practitioners who want to quantify the effect of some regressor x on the response probability, other things being equal. For the logit model in (1), the partial effect of covariate x_{itk} for i at time t on the probability of $y_{it} = 1$ can be written, depending on the typology of covariate, as

$$m_{itk}(\alpha_i, \boldsymbol{\beta}, \boldsymbol{x}_{it}) = \begin{cases} p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it})[1 - p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it})]\beta_k, & x_{itk} \text{ continuous} \\ p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it,-k}, x_{itk} = 1) - \\ \qquad p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it,-k}, x_{itk} = 0), & x_{itk} \text{ discrete} \end{cases}$$

4

where $\boldsymbol{x}_{it,-k}$ denotes the subvector of all covariates but x_{itk}. The true APE of the k-th covariate can then be obtained by simply taking the expected value of $f_{itk}(\alpha_i, \boldsymbol{\beta}, \boldsymbol{x}_{it})$ with respect to $\boldsymbol{x}_{it}$:

$$\mu_{k0} \equiv \mathrm{E}_{nT}\left[m_{itk}(\alpha_{i0}, \boldsymbol{\beta}_0, \boldsymbol{x}_{it})\right],$$

where $\mu_{k0} \equiv \mu_k(\alpha_{i0}, \boldsymbol{\beta}_0)$. An estimator of μ_{k0} can be obtained by plugging in the ML estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}_i(\hat{\boldsymbol{\beta}})$, so that

$$\hat{\mu}_k = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} m_{itk}(\hat{\alpha}_i(\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}, \boldsymbol{x}_{it}). \tag{2}$$

It is now clear that, with T fixed, this estimator is plagued by two sources of asymptotic bias: the first stems from the estimation error introduced by $\hat{\alpha}_i(\boldsymbol{\beta})$, used instead of $\alpha_i(\boldsymbol{\beta})$; the second is a result of using the asymptotically biased estimator $\hat{\boldsymbol{\beta}}$. Dhaene and Jochmans (2015) show that the combined asymptotic bias is

$$\operatorname*{plim}_{n\to\infty} \hat{\mu}_k = \mu_{k0} + \frac{D + E}{T} + O(T^{-2}), \tag{3}$$

where, specifically, D is the bias the generates from using $\hat{\alpha}_i(\boldsymbol{\beta})$ instead of α_{i0}, whereas E is the bias from plugging in $\hat{\boldsymbol{\beta}}$, instead if using $\boldsymbol{\beta}_0$. Dhaene and Jochmans (2015) provide explicit expressions for D and E, based on the derivations by Fernández-Val (2009). Notice that, even if a fixed–T consistent estimator of $\boldsymbol{\beta}_0$ was available, the asymptotic bias of the APE estimator would still be of order $O(T^{-1})$ and equal to D/T.

3   Estimation of average partial effects

The previous section clarifies that a bias corrected estimators of μ_{k0} must take into account the two sources of asymptotic bias combined in (3). In the following, we first illustrate the proposed methodology, which combines the consistent CML estimator of $\boldsymbol{\beta}_0$ and a bias corrected estimator of α_{i0}. We then turn to the dynamic logit, for which the proposed procedure is based on a PCML estimator. Finally, we briefly review the existing strategies with special attention to the jackknife procedure, that represent the benchmark against which to compare the finite sample performance of the proposed estimator.

3.1   Proposed methodology

The proposed two-step strategy is based on removing the two sources of bias in (3) by i) using the fixed-T consistent Conditional Maximum Likelihood (CML) estimator of $\boldsymbol{\beta}_0$, $\tilde{\boldsymbol{\beta}}$ instead of the ML estimator $\hat{\boldsymbol{\beta}}$ and ii) reducing the order of bias of $\hat{\alpha}_i(\tilde{\boldsymbol{\beta}})$ from $O(T^{-1})$ to $O(T^{-2})$.

3.1.1 Two step estimation

The first step consists estimating by CML the structural parameters of the logit model in (1). Taking the the individual intercept α_i as given, The joint probability of the response configuration $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ conditional on $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})$ can be written as

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

It can be shown that the *total score* $y_{i+} = \sum_t y_{it}$ is a sufficient statistic for the individual intercepts α_i (Andersen, 1970). The joint probability of $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})$ conditional on y_{i+} does not depend on α_i and can therefore be written as

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}) = \frac{\exp\left[\left(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp\left[\left(\sum_{t=1}^T z_t\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]}, \tag{4}$$

where the denominator is the sum over all the response configuration $\boldsymbol{z}$ such that $z_+ = y_{i+}$ and where the individual intercepts α_i have been canceled out. The log-likelihood function is

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

where the indicator function $\mathbf{I}(\cdot)$ takes into account that observations with total score y_{i+} equal to 0 or T do not contribute to the log-likelihood and $p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+})$ is defined in (4). The above function can be maximized with respect to $\boldsymbol{\beta}$ by a Newton-Raphson algorithm using standard results on the regular exponential family (Barndorff-Nielsen, 1978), so as to obtain the CML estimator $\tilde{\boldsymbol{\beta}}$, which is $\sqrt{n}$ consistent and asymptotically normal with fixed–T (see Andersen, 1970; Chamberlain, 1980, for details). Therefore, if plugged into the APE formulation (2) instead of the ML estimator $\hat{\boldsymbol{\beta}}$, the E component of the bias in (3) is removed. Alternatively, Hahn and Newey (2004) suggest using a biased corrected version of $\hat{\boldsymbol{\beta}}$ for which, however, T has to grow faster than n for its asymptotic distribution to be centered at its probability limit.

The second step deals with obtaining estimates of the individual intercepts α_i, which are not directly available as they have been canceled out by conditioning on the total score. One strategy would be to obtain the ML estimates of α_i, for those subjects such that $0 < y_{i+} < T$, by maximizing the individual $\sum_t \log p_{\tilde{\boldsymbol{\beta}}}(y_{it}|\alpha_i, \boldsymbol{x}_{it})$ where $p_{\tilde{\boldsymbol{\beta}}}(y_{it}|\alpha_i, \boldsymbol{x}_{it})$ is the logit model probability in (1) evaluated at the CML estimate $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$. This strategy has been considered by Stammann et al. (2016). However, even if $\boldsymbol{\beta}$ is fixed at some $\sqrt{n}$-consistent estimate, the bias of the ML estimator of α_{i0} will still be of order $O(T^{-1})$

because $\hat{\alpha}_i(\tilde{\boldsymbol{\beta}}) \overset{p}{\to} \alpha_{i0}$ only as $T \to \infty$.

Our strategy is based on the proposal by Firth (1993), who shows that, for any ML estimator $\hat{\boldsymbol{\psi}}$ with bias of order $O(h^{-1})$, where h is the number of observations, the score function $U(\boldsymbol{\psi})$ can be modified as

$$U^*(\boldsymbol{\psi}) = U(\boldsymbol{\psi}) + \frac{1}{2}tr\left[I(\boldsymbol{\psi})^{-1}\frac{\partial I(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'}\right],$$

where $I(\boldsymbol{\psi})$ is the Fisher Information matrix, so that the solution to the above estimating equation is an ML estimator with bias $O(h^{-2})$.[2] We therefore obtain $\tilde{\alpha}_i(\tilde{\boldsymbol{\beta}})$ as the solution to

$$U^*(\alpha_i) = U(\alpha_i) + \frac{1}{2I(\alpha_i)}\frac{\partial I(\alpha_i)}{\partial \alpha_i} =$$

$$\sum_{t=1}^{T}(y_{it} - \tilde{r}_{it}) + \frac{\sum_{t=1}^{T}\tilde{r}_{it}(1 - \tilde{r}_{it})(1 - 2\tilde{r}_{it})}{2\left[\sum_{t=1}^{T}\tilde{r}_{it}(1 - \tilde{r}_{it})\right]^2},$$

where $\tilde{r}_{it} = \exp(\alpha_i + \boldsymbol{x}_{it}'\tilde{\boldsymbol{\beta}})[1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\tilde{\boldsymbol{\beta}})]$. The resulting estimator of the individual intercept $\tilde{\alpha}_i$ will depend on $\tilde{\boldsymbol{\beta}}$, which we write as $\tilde{\alpha}_i(\tilde{\boldsymbol{\beta}})$. The APEs can then be obtained by simply replacing the ML estimators in (2), that is

$$\tilde{\mu}_k = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}m_{itk}(\tilde{\alpha}_i(\tilde{\boldsymbol{\beta}}), \tilde{\boldsymbol{\beta}}, \boldsymbol{x}_{it}).$$

3.1.2 Standard errors

In order to derive an expression for the standard errors of the APEs $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_K)'$ we need to account for the use of the estimated parameters $\tilde{\boldsymbol{\beta}}$ in the first step. We rely on the Generalized Method of Moments (GMM) approach by Hansen (1982) and also implemented by Bartolucci and Nigro (2012) for the Quadratic Exponential model. It consists in presenting the proposed multi-step procedure as the solution of the system of estimating equations

$$\boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \boldsymbol{0},$$

where

$$\boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \sum_{i=1}^{n}\mathbf{I}(0 < y_{i+} < T)\boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{\mu}),$$

---

[2]This procedure is also mentioned by Hahn and Newey (2004) and Fernández-Val (2009) relatively to the estimation of $\boldsymbol{\beta}$

$$\boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{\mu}) = \begin{pmatrix} \boldsymbol{\nabla}_{\boldsymbol{\beta}} \ell_i(\boldsymbol{\beta}) \\ \boldsymbol{\nabla}_{\mu_1} g_i(\boldsymbol{\beta}, \mu_1) \\ \vdots \\ \boldsymbol{\nabla}_{\mu_K} g_i(\boldsymbol{\beta}, \mu_K) \end{pmatrix}, \tag{5}$$

and

$$g_i(\boldsymbol{\beta}, \mu_k) = \frac{1}{T} \sum_t \left[m_{itk}(\alpha_i(\boldsymbol{\beta}), \boldsymbol{\beta}, \boldsymbol{x}_{it}) - \mu_k \right]^2, \ k = 1, \dots, K.$$

The asymptotic variance of $(\tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\mu}}')'$ is then

$$\boldsymbol{W}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}}) = \boldsymbol{H}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}})^{-1} \boldsymbol{S}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}}) [\boldsymbol{H}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}})^{-1}]', \tag{6}$$

where

$$\boldsymbol{S}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}}) = \sum_i \boldsymbol{\mathrm{I}}(0 < y_{i+} < T) \boldsymbol{f}_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}}) \boldsymbol{f}_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}})',$$

$$\boldsymbol{H}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}}) = \sum_i \boldsymbol{\mathrm{I}}(0 < y_{i+} < T) \boldsymbol{H}_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}}),$$

and

$$\boldsymbol{H}_i(\boldsymbol{\beta}, \boldsymbol{\mu}) = \begin{pmatrix} \boldsymbol{\nabla}_{\boldsymbol{\beta}\boldsymbol{\beta}} \ell_i(\boldsymbol{\beta}) & \boldsymbol{O} \\ \boldsymbol{\nabla}_{\boldsymbol{\mu}\boldsymbol{\beta}} \boldsymbol{g}_i(\boldsymbol{\beta}, \boldsymbol{\mu}) & \boldsymbol{\nabla}_{\boldsymbol{\mu}\boldsymbol{\mu}} \boldsymbol{g}_i(\boldsymbol{\beta}, \boldsymbol{\mu}) \end{pmatrix}, \tag{7}$$

is the derivative of $\boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{\mu})$ with respect to $(\boldsymbol{\beta}, \boldsymbol{\mu})$, where $\boldsymbol{O}$ denotes a $K \times K$ matrix of zeros and $\boldsymbol{g}_i(\boldsymbol{\beta}, \boldsymbol{\mu})$ collects $g_i(\boldsymbol{\beta}, \mu_k)$, for $k = 1, \dots, K$. Expressions for the derivatives in (5) are

$$\nabla_{\boldsymbol{\beta}} \ell_i(\boldsymbol{\beta}) = \sum_{t=1}^{T} y_{it} \boldsymbol{x}_{it} - \sum_{\boldsymbol{z}: z_+ = y_{i+}} \left(p(\boldsymbol{z}|\boldsymbol{X}_i, y_{i+}) \sum_{t=1}^{T} z_t \boldsymbol{x}_{it} \right),$$

and

$$\nabla_{\mu_k} g_i(\boldsymbol{\beta}, \mu_k) = -\frac{2}{T} \sum_{t=1}^{T} \left[m_{itk}(\alpha_i(\boldsymbol{\beta}), \boldsymbol{\beta}, \boldsymbol{x}_{it}) - \mu_k \right].$$

The second derivatives in (7) are

$$\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}} \ell_i(\boldsymbol{\beta}) = \sum_{\boldsymbol{z}: z_+ = y_{i+}} p(\boldsymbol{z}|\boldsymbol{X}_i, y_{i+}) \boldsymbol{e}(\boldsymbol{z}, \boldsymbol{X}_i) \boldsymbol{e}(\boldsymbol{z}, \boldsymbol{X}_i)',$$

where

$$\boldsymbol{e}(\boldsymbol{z}, \boldsymbol{X}_i) = \sum_{t=1}^{T} z_t \boldsymbol{x}_{it} - \sum_{\boldsymbol{z}: z_+ = y_{i+}} \left(p(\boldsymbol{z}|\boldsymbol{X}_i, y_{i+}) \sum_{t=1}^{T} z_t \boldsymbol{x}_{it} \right),$$

and $\boldsymbol{\nabla}_{\boldsymbol{\mu}\boldsymbol{\mu}} \boldsymbol{g}_i(\boldsymbol{\beta}, \boldsymbol{\mu})$ is a $K \times K$ diagonal matrix with element 2. Finally, for the computation of the block $\boldsymbol{\nabla}_{\boldsymbol{\mu}\boldsymbol{\beta}} g_i(\boldsymbol{\beta}, \boldsymbol{\mu})$ we rely on numerical differentiation. Once the matrix in (6) is computed, the standard errors for the APEs $\tilde{\boldsymbol{\mu}}$ may be obtained by taking the square root of the elements in the main diagonal of the lower right submatrix of $\boldsymbol{W}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}})$.

3.1.3 The dynamic logit model

The method proposed to obtain the APE for the logit model cannot be applied directly to the dynamic logit (Hsiao, 2005). For the dynamic logit model, the conditional probability of y_{it} being equal to 1 is

$$p(y_{it}|\boldsymbol{x}_{it}, y_{i,t-1}; \eta_i, \boldsymbol{\delta}, \gamma) = \frac{\exp\left[y_{it}(\eta_i + \boldsymbol{x}'_{it}\boldsymbol{\delta} + y_{i,t-1}\gamma)\right]}{1 + \exp(\eta_i + \boldsymbol{x}'_{it}\boldsymbol{\delta} + y_{i,t-1}\gamma)}, \tag{8}$$

where γ is the regression coefficient for the lagged response variable that measures the true state dependence. Plugging the CML estimator of $\boldsymbol{\delta}$ and γ in the APE formulation is not viable in this case because the total score is no longer a sufficient statistic for the incidental parameters if the lag of the dependent variable is included among the model covariates. Conditioning in sufficient statistics eliminates the incidental parameters only in the in the special case of $T = 3$ and no other explanatory variables (Chamberlain, 1985). Honoré and Kyriazidou (2000) extend this approach to include explanatory variables and parameters can be estimated by CML on the basis of a weighted conditional log-likelihood. However, time effects cannot be included in the model specification and the estimator's rate of convergence to the true parameter value is slower than $\sqrt{n}$. This is overcome by Bartolucci and Nigro (2010), who propose a Quadratic Exponential (QE) formulation (Cox, 1972) to model dynamic binary panel data, that has the advantage of admitting sufficient statistics for the individual intercepts.

Bartolucci and Nigro (2012) propose a QE model, that approximates more closely the dynamic logit model, the parameters of which can easily be estimated by PCML. Under the approximating model, each y_{i+} is a sufficient statistic for the fixed effect η_i. By conditioning on the total score, the joint probability of $\boldsymbol{y}_i$ becomes:

$$p^*(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\delta} - \sum_t \bar{q}_{it}y_{i,t-1}\gamma + y_{i*}\gamma)}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\delta} - \sum_t \bar{q}_{it}z_{i,t-1}\gamma + z_{i*}\gamma)}, \tag{9}$$

where $y_{i*} = \sum_t y_{i,t-1}y_{it}$, and $z_{i*} = y_{i0}z_1 + \sum_{t>1} z_{t-1}z_t$. Moreover, $\bar{q}_{it}$ is a function of given values of $\boldsymbol{\delta}$ and η_i, resulting from a first-order Taylor-series expansion of the log-likelihood based on (8) around $\boldsymbol{\delta} = \bar{\boldsymbol{\delta}}$ and $\eta_i = \bar{\eta}_i$, $i = 1, \ldots, n$, and $\gamma = 0$ (see Bartolucci and Nigro, 2012, for details). The expression for $\bar{q}_{it}$ is then

$$\bar{q}_{it} = \frac{\exp(\bar{\eta}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\delta}})}{\left[1 + \exp(\bar{\eta}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\delta}})\right]}.$$

Expressions for the partial effects and APEs are derived in the same way as for the static logit model. Let $\boldsymbol{w}_{it} = (\boldsymbol{x}'_{it}, y_{it-1})'$ collect the $K + 1$ model covariates. Based on (8), the partial effect of covariate w_{itk} for i at time t on the probability of $y_{it} = 1$ can be

written as

$$
v_{itk}(\eta_i, \boldsymbol{\theta}, \boldsymbol{w}_{it}) = \begin{cases} p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it}) \left[1 - p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it})\right]\delta_k, & w_{itk} \text{ continuous} \\[2mm] \begin{aligned} &p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it,-k}, w_{itk} = 1)- \\ &\quad p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it,-k}, w_{itk} = 0), \end{aligned} & w_{itk} \text{ discrete} \end{cases}
$$

where $\boldsymbol{w}_{it,-k}$ again denotes the the vector $\boldsymbol{w}_{it}$ excluding w_{itk}, and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \gamma)'$. Notice that this function does not depend on $\bar{\boldsymbol{\delta}}$, since the probability in (8) does not depend on $\bar{q}_{it}$. The APE of the k-th covariate can then be obtained by taking the expected value of $v_{itk}(\eta_i, \boldsymbol{\theta}, \boldsymbol{w}_{it})$ with respect to $\boldsymbol{w}_{it}$ and evaluated in η_{i0}, $\boldsymbol{\theta}_0$, and $\boldsymbol{w}_{it}$ can be written as

$$
\nu_{k0} \equiv \mathrm{E}_{nT}\left[v_{itk}(\eta_{i0}, \boldsymbol{\theta}_0, \boldsymbol{w}_{it})\right],
$$

where $\nu_{k0} \equiv \nu_k(\eta_{i0}, \boldsymbol{\theta}_0)$.

As for the static logit model, the estimation of ν_{k0} requires an estimate of η_i, which we obtain in the same manner as in the second step in Section 3.1.1. Here, however, the CML estimation of $\boldsymbol{\theta}$ based on (9) relies on a preliminary step in order to obtain $\bar{q}_{it}$ and the estimation of APEs is thus based on a three-step procedure.

In the first step, a preliminary estimate of $\bar{\boldsymbol{\delta}}$ is obtained by maximizing the conditional log-likelihood

$$
\ell(\bar{\boldsymbol{\delta}}) = \sum_i \mathbf{I}(0 < y_{i+} < T)\ell_i(\bar{\boldsymbol{\delta}}),
$$

where

$$
\ell_i = \log \frac{\exp\left[\left(\sum_t y_{it}\boldsymbol{x}_{it}\right)' \bar{\boldsymbol{\delta}}\right]}{\displaystyle\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp\left[\left(\sum_t z_t\boldsymbol{x}_{it}\right)' \bar{\boldsymbol{\delta}}\right]},
$$

which is the same conditional log-likelihood of the static logit model and may be maximized by a standard Newton-Raphson algorithm. We denote the resulting CML estimator by $\check{\boldsymbol{\delta}}$. The estimate $\check{\eta}_i$ is then computed by maximizing the individual log-likelihood

$$
\ell_i(\bar{\eta}_i) = \sum_t \log \frac{\exp\left[y_{it}(\bar{\eta}_i + \boldsymbol{x}'_{it}\check{\boldsymbol{\delta}})\right]}{1 + \exp(\bar{\eta}_i + \boldsymbol{x}'_{it}\check{\boldsymbol{\delta}})},
$$

where $\check{\boldsymbol{\delta}}$ is fixed. The probability $\bar{q}_{it}$ in (9) can the be estimated by $\check{q}_{it} = \exp(\check{\eta}_i + \boldsymbol{x}'_{it}\check{\boldsymbol{\delta}})/\left[1 + \exp(\check{\eta}_i + \boldsymbol{x}'_{it}\check{\boldsymbol{\delta}})\right]$.

In the second step, we estimate $\boldsymbol{\theta}$ by maximizing the following conditional log-likelihood

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

where $p^*_{\check{\boldsymbol{q}}_i}(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+})$ is the joint probability in (9) evaluated at $\check{\boldsymbol{q}}_i = (\check{q}_{i1}, \ldots, \check{q}_{iT})'$. The above function can be easily maximized with respect to $\boldsymbol{\theta}$ by the Newton-Raphson

algorithm, so as to obtain the PCML estimator $\tilde{\boldsymbol{\theta}}$, which is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$ only if $\gamma_0 = 0$, representing the special case in which the QE model corresponds to the dynamic logit model.[3]. Nonetheless, Bartolucci and Nigro (2012) show that the PCML estimator has a limited bias in finite sample even in presence of non negligible state dependence.

Finally, in step three, we obtain an estimate of η_i as a solution to the modified score function by Firth (1993), that can be written as

$$U^*(\eta_i) = U(\eta_i) + \frac{1}{2I(\eta_i)} \frac{\partial I(\eta_i)}{\partial \eta_i} =$$

$$\sum_{t=1}^{T} (y_{it} - \tilde{s}_{it}) + \frac{\sum_{t=1}^{T} \tilde{s}_{it}(1 - \tilde{s}_{it})(1 - 2\tilde{s}_{it})}{2\left[\sum_{t=1}^{T} \tilde{s}_{it}(1 - \tilde{s}_{it})\right]^2},$$

where $\tilde{s}_{it} = \exp(\eta_i + \boldsymbol{w}_{it}'\tilde{\boldsymbol{\theta}})/[1 + \exp(\eta_i + \boldsymbol{w}_{it}'\tilde{\boldsymbol{\theta}})]$. The resulting estimator of the individual intercept $\tilde{\eta}_i$ depends on $\tilde{\boldsymbol{\theta}}$, which we write as $\tilde{\eta}_i(\tilde{\boldsymbol{\theta}})$.

The APEs can then be estimated by plugging $\tilde{\eta}_i(\tilde{\boldsymbol{\theta}})$ and $\tilde{\boldsymbol{\theta}}$ in the APE formulation, so as to obtain

$$\tilde{\nu}_k = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} v_{itk}(\tilde{\eta}_i(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\theta}}, \boldsymbol{w}_{it}).$$

Standard errors for $\tilde{\nu}_k$ can be obtained exactly in the same way as illustrated in Section 3.1.2 with the appropriate change of notation.

3.2 Alternative strategies

Along the lines of the proposals put forward to remove the bias from the ML estimator, the available bias reduction techniques for the estimation of APEs for fixed-effects binary choice models are mainly based on either analytical or jackknife bias corrections.[4]

Analytical bias corrections amount to deriving the two sources of bias D and E in (3) in order to evaluate their sample counterparts and find a bias corrected estimator $\hat{\mu}_k^c = \hat{\mu}_k - (\hat{D} + \hat{E})/T$. The asymptotic bias arising from plugging in $\hat{\alpha}_i(\boldsymbol{\beta})$ can be written as

$$D = \sum_{j=0}^{\infty} \mathrm{E}_{nT}\left[\frac{\partial \mu_k(\alpha_{i0}, \boldsymbol{\beta}_0)}{\partial \alpha_{i0}} \tau_{it-j}\right] + \mathrm{E}_{nT}\left[\frac{\partial \mu_k(\alpha_{i0}, \boldsymbol{\beta}_0)}{\partial \alpha_{i0}} \xi_i\right] + \frac{1}{2}\mathrm{E}_{nT}\left[\frac{\partial^2 \mu_k(\alpha_{i0}, \boldsymbol{\beta}_0)}{\partial \alpha_{i0}^2} \sigma_i^2\right],$$

where expressions for ψ_{is}, ξ_i, and σ_i^2 for panel binary choice models are given in Fernández-

---

[3]The correspondence refers to the log-odds ratio. This is clarified by Theorem 1 in Bartolucci and Nigro (2012).

[4]In the following discussion, we will use the notation for the static logit model, unless required otherwise. Nonetheless, everything that follows can be generalized to the dynamic logit model.

Val (2009), [5] whose derivations are based on general results for static (Hahn and Newey, 2004) and dynamic (Hahn and Kuersteiner, 2011) nonlinear panel data models.[6] The asymptotic bias of $\hat{\mu}_k$ deriving from $\hat{\boldsymbol{\beta}}$ can be written as

$$E = \mathrm{E}_{nT} \left[\frac{\partial \mu_k(\alpha_{i0}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0'} \right] B,$$

where B is the leading term of the large-T expansion for the asymptotic bias of $\hat{\boldsymbol{\beta}}$ (see Fernández-Val, 2009). Notice that if a bias corrected estimator of $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}^c$, was used instead of $\hat{\beta}$ to evaluate $\hat{\mu}_k$, then only the D term would have to be removed in order to obtain the bias reduction, as suggested by Hahn and Newey (2004). For the expressions as well as for further details we refer the reader to Hahn and Newey (2004), Fernández-Val (2009), and Hahn and Kuersteiner (2011).

An alternative bias correction method for the APE estimator relies on the panel jack-knife. A general procedure for nonlinear static panel data models in proposed by Hahn and Newey (2004). Let $\hat{\boldsymbol{\beta}}^{(t)}$ and $\hat{\alpha}_i^{(t)}(\hat{\boldsymbol{\beta}}^{(t)})$ be the ML estimators with the t-th observation excluded for each subject. Then the jackknife corrected estimator for the APE is

$$\hat{\mu}_k^c = T\hat{\mu}_k - \frac{T-1}{T} \sum_{t=1}^{T} \mu_k \left(\hat{\alpha}_i^{(t)}(\hat{\boldsymbol{\beta}}^{(t)}), \hat{\boldsymbol{\beta}}^{(t)} \right).$$

If the set of model covariates includes the lag of explanatory variables, then leaving out one of the t observations at the time becomes unsuitable. Instead, a block of consecutive observations has to be considered so as to preserve the dynamic structure of the data. The so-called split panel jackknife estimator was proposed by Dhaene and Jochmans (2015). A simple version of the estimator is the half-panel jackknife, which is based on splitting the panel into two half-panels, also non-overlapping if T is even and $T \geq 6$, and with $T/2$ time periods. Denote the set of half-panels as

$$S = \{S_1, S_2\}, \quad S_1 = \{1, \ldots, T/2\}, S_2 = \{T/2 + 1, \ldots, T\},$$

then the half-panel jackknife estimator of the APE is

$$\hat{\nu}_k^{1/2} = 2\hat{\nu}_k - \frac{1}{2} \left(\bar{\nu}_k^{S_1} + \bar{\nu}_k^{S_2} \right),$$

---

[5]The term ξ_i is denoted by β_i and the term τ_{it} by ψ_{it} in Fernández-Val (2009).

[6]The expression for D is a function of the asymptotic bias and variance components of $\hat{\alpha}_i(\boldsymbol{\beta})$, that is

$$\hat{\alpha}_i(\boldsymbol{\beta}) = \alpha_{i0} + \frac{\xi_i}{T} + \frac{1}{T} \sum_{t=1}^{T} \tau_{it} + o_p \left(\frac{1}{T} \right),$$

where $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_{it} \xrightarrow{d} N(0, \sigma_i^2)$ (see Fernández-Val, 2009; Dhaene and Jochmans, 2015, for details).

where $\bar{\nu}_k^{S_1}$ and $\bar{\nu}_k^{S_2}$ are the plug-in estimators evaluated at the ML estimators of $\eta_i(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$ obtained using the observations in subpanels S_1 and S_2, respectively. Dhaene and Jochmans (2015) also illustrate generalized versions of the half-panel jackknife to deal with odd T and overlapping subpanels, as well as an alternative jackknife estimator based on the split-panel log-likelihood correction.

As well as for bias corrected fixed-effects estimator, the analytical and jackknife corrections reduce the order of the bias of the APE estimator $\hat{\mu}_k$ (or $\hat{\nu}_k$) from $O(T^{-1})$ to $O(T^{-2})$, which is also the case of the proposed method. It is worth stressing, however, that the APE estimators discussed in this section still depend on some estimator of $\boldsymbol{\beta}_0$, corrected by either analytical or jackknife procedures. In both cases, it can be shown that, in order for $\hat{\boldsymbol{\beta}}^c$ to have correct confidence intervals, T has to grow faster than $n^{1/3}$ (Hahn and Newey, 2004). If this is not the case, then the asymptotic distribution $\hat{\boldsymbol{\beta}}^c$ will not be centered in $\boldsymbol{\beta}_0$ and this source of distortion will affect the asymptotic distribution of $\hat{\mu}_k^c$, since E is not correctly removed. This is in contrast with the procedure here proposed, which is based on a fixed-T consistent estimator of $\boldsymbol{\beta}_0$.

4 Simulation study

In the following we illustrate the design and discuss the results of the simulation studies aimed at assessing the finite sample performance of the estimators of the APEs for the static and dynamic logit models. We keep the analyses separate for the two models, as we base the two studies on different simulation designs.

4.1 Static logit

The simulation design for the static logit model is based on the one adopted by Hahn and Newey (2004), except that we consider logit rather than normal error terms. The data are generated as

$$y_{it} = \mathbf{I}(\alpha_i + x_{it}\beta + \varepsilon_{it} > 0), \quad i = 1, \dots, n, \ t = 1, \dots, T,$$

with $\alpha_i \sim N(0,1)$, ε_{it} follow a standard logistic distribution, and

$$x_{it} = t/10 + x_{i,t-1}/2 + u_{it},$$

where $u_{it} \sim U[-0.5, 0.5]$ and $x_{i0} = u_{i0}$. We consider different scenarios according to the values of n and T and we set $n = 100, 500, 1000$, $T = 4, 8, 12$. Hahn and Newey (2004) considered only $n = 100$ and $T = 4, 8$. The coefficient β is equal to 1 across all the scenarios and the number of replications is 1000.

Table 1 reports the simulation results for each scenario. We compare the finite sample performance of the proposed APE estimator (denoted by CML - ML Firth) with i) an estimator for μ based on plugging in the CML estimator of β and the ML estimator of $\alpha_i(\tilde{\beta})$ that is not bias corrected by using Firth (1993)'s modified score (CML - ML) and with ii) Hahn and Newey (2004)'s jackknife bias corrected estimator, briefly recalled in Section 3.2 (Jackknife).[7] For each scenario, we report the mean and the median of the ratio $\tilde{\mu}/\mu$, the standard deviation of $\tilde{\mu}$, the interval coverage at the confidence level 90% and 95%, and the mean ratio between the estimator standard error and standard deviation.[8]

From Table 1, it emerges that the proposed estimator (CML - ML Firth) has a good finite sample performance with both small n and T, and even when n is much larger than T. This result suggests that Firth (1993)'s correction is working nicely in removing the bias component D in (3), as also testified by the rather poor performance of the CML - ML estimator, based on $\hat{\alpha}_i(\tilde{\beta})$, which still exhibits some bias even with $T = 12$.

The proposed procedure and the jackknife estimator exhibit a similar behavior in the scenario with $n = 100$, which is also in line with the results reported by Hahn and Newey (2004) for the probit model, with the exception of scenario with $T = 4$, where CML - ML Firth actually performs better than the jackknife. It is worth stressing that while the CML estimator of β is $\sqrt{n}$-consistent, the jackknife bias corrected slope estimator requires T to grow faster than $n^{1/3}$. This reflects indirectly on the estimator of the partial effect, since a bias corrected or consistent estimator of β takes care of removing the term E from (3). It is therefore clear why the the proposed estimator performs better than the jackknife with, say, a sample size as large as $n = 1000$, unless $T = 12$. It is also worth noticing that, throughout the scenarios, the proposed estimator has a good interval coverage, with the percentage attaining the nominal confidence level as T grows.

4.2   Dynamic logit

For the static logit model, the simulation design is similar to that by Dhaene and Jochmans (2015), where again we consider a logit rather than a normal distribution for the error terms. The data generating process is as follows

$$y_{it} = \mathbf{I}(\eta_i + y_{i,t-1}\gamma + \upsilon_{it} > 0), \quad i = 1, \dots, n, \ t = 1, \dots, T,$$

---

[7]We do not report simulation results for analytically bias corrected estimators of the type $\hat{\mu}^c$ discussed in Section 3.2. Based on the same simulation design and for the scenarios with $n = 100$, results are reported for the probit model by Fernández-Val (2009), who shows that the finite sample performance of these estimator is quite similar to that of the jackknife estimator.

[8]For the CML-ML, the standard error is computed using the procedure in Section 3.1.2, whereas jackknife standard errors are computed for Hahn and Newey (2004)'s estimator.

Table 1: Simulation results for $\tilde{\mu}$, static logit model

n	T		Mean ratio	Median ratio	SD	Confidence 90%	95%	SE/SD
100	4							
		CML - ML Firth	1.036	1.032	0.056	0.974	0.990	1.361
		CML - ML	0.974	0.972	0.053	0.976	0.989	1.359
		Jackknife	1.287	1.275	0.074	0.847	0.901	1.328
100	8							
		CML - ML Firth	1.004	1.001	0.026	0.924	0.968	1.085
		CML - ML	0.953	0.950	0.025	0.915	0.964	1.087
		Jackknife	1.079	1.074	0.028	0.868	0.912	1.147
100	12							
		CML - ML Firth	0.964	0.966	0.015	0.889	0.929	1.019
		CML - ML	0.931	0.934	0.015	0.812	0.897	1.021
		Jackknife	1.009	1.013	0.016	0.950	0.971	1.325
500	4							
		CML - ML Firth	1.040	1.040	0.024	0.980	0.992	1.419
		CML - ML	0.977	0.976	0.023	0.972	0.996	1.418
		Jackknife	1.301	1.306	0.032	0.666	0.742	1.316
500	8							
		CML - ML Firth	1.001	1.001	0.012	0.939	0.973	1.092
		CML - ML	0.950	0.950	0.011	0.862	0.928	1.092
		Jackknife	1.078	1.077	0.013	0.793	0.868	1.230
500	12							
		CML - ML Firth	1.013	1.013	0.007	0.909	0.958	1.022
		CML - ML	0.968	0.969	0.007	0.838	0.912	1.025
		Jackknife	1.048	1.049	0.007	0.979	0.995	1.769
1000	4							
		CML - ML Firth	1.042	1.040	0.018	0.972	0.987	1.373
		CML - ML	0.978	0.978	0.017	0.973	0.995	1.373
		Jackknife	1.306	1.304	0.023	0.513	0.602	1.313
1000	8							
		CML - ML Firth	1.006	1.007	0.008	0.929	0.974	1.074
		CML - ML	0.954	0.956	0.008	0.799	0.879	1.077
		Jackknife	1.083	1.083	0.009	0.684	0.797	1.311
1000	12							
		CML - ML Firth	1.013	1.013	0.005	0.887	0.935	0.960
		CML - ML	0.969	0.968	0.005	0.750	0.851	0.960
		Jackknife	1.049	1.048	0.005	0.996	1.000	2.158

Notes: 1000 replications. CML-ML Firth denotes the proposed estimator; CML-ML denotes the estimator of the APE based on the CML estimate of β and the uncorrected estimated of $\alpha_i(\tilde{\beta})$; Jackknife denotes Hahn and Newey (2004)'s jackknife bias corrected estimator.

with $\eta_i \sim N(0, 1)$, v_{it} follow a standard logistic distribution, and the initial condition is $y_{i0} = \mathbf{I}(\eta_i + +v_{i0} > 0)$. We consider the same scenarios as for the static logit model, that are $n = 100, 500, 1000$, $T = 4, 8, 12$. The coefficient γ is equal to 0.5 across all the scenarios and the number of replications is 1000.

Table 2 reports the simulation results where we compare the finite sample performance of the proposed APE estimator (PCML - ML Firth) with the estimator for ν based on the PCML estimator of γ and the uncorrected ML estimator $\hat{\eta}_i(\tilde{\gamma})$ (PCML - ML) and with Dhaene and Jochmans (2015)'s half-panel jackknife bias corrected estimator (Jackknife) illustrated in Section 3.2. [9] Again we report the mean and the median of the ratio $\tilde{\nu}/\nu$, the standard deviation of $\tilde{\nu}$, the interval coverage at the confidence level 90% and 95%, and the mean ratio between the estimator standard error and standard deviation.[10]

Table 2 reports the simulation results, from which it emerges that the proposed estimator exhibits a better finite sample performance than the half-panel jackknife estimator. Notice also that the jackknife estimator for the dynamic logit model cannot be computed for $T < 6$, as clarified in Section 3.2. Clearly the finite sample performance of the proposed estimator deteriorates with respect to that of CML - ML Firth, as per the effect of the PCML estimator, which is not consistent when $\gamma \neq 0$. Still, it can be noticed that the bias reduces rather quickly as T grows, as an effect of the use of the corrected $\tilde{\eta}_i(\tilde{\gamma})$ in the computation of $\tilde{\nu}$, and that the proposed estimator represents a substantial improvement upon the half-panel jackknife in scenarios where $T = 8, 12$.

5   Empirical application

We apply our proposed formulation to the problem of estimating the labor supply of married women. The same empirical application is considered by Fernández-Val (2009) and Dhaene and Jochmans (2015). The sample is drawn from the Panel Study of Income Dynamics (PSID), that consists of $n = 1,908$ married women between 19 and 59 years of age in 1980, followed for $T = 7$ time occasions, from 1979 to 1985. We specify a static logit model for the probability of being employed at time t, conditional on the number of children of a certain age in the family, namely the number of kids between 0 and 2 years old, between 3 and 5, and between 6 and 17, on the husband's income, and on the woman's age and age squared. We also specify a dynamic logit model, that is we include lagged participation in the set of model covariates.

---

[9]In the Monte Carlo study by Dhaene and Jochmans (2015), the APE estimator is computed for the covariate x_{it}, which is associated with $\delta = 0$ in the current design. We instead investigate the finite sample performance of the APE estimator for the effect of $y_{i,t-1}$.

[10]For the PCML-ML, the standard error is computed using the procedure in Section 3.1.2, whereas the standard error estimates for the half-panel jackknife are based on the cross-sectional variance of the within-group average effects, as suggested by Dhaene and Jochmans (2015).

Table 2: Simulation results for $\tilde{\nu}$, dynamic logit model

n	T		Mean ratio	Median ratio	SD	Confidence 90%	95%	SE/SD
100	4							
		PCML - ML Firth	0.885	0.874	0.066	0.979	0.995	1.478
		PCML - ML	0.837	0.820	0.062	0.980	0.997	1.465
		Jackknife						
100	8							
		PCML - ML Firth	0.945	0.947	0.041	0.924	0.963	1.119
		PCML - ML	0.893	0.898	0.039	0.912	0.948	1.119
		Jackknife	0.775	0.756	0.043	0.864	0.898	1.254
100	12							
		PCML - ML Firth	0.941	0.939	0.031	0.910	0.957	1.068
		PCML - ML	0.900	0.900	0.030	0.897	0.954	1.068
		Jackknife	0.877	0.876	0.033	0.853	0.884	1.194
500	4							
		PCML - ML Firth	0.896	0.902	0.030	0.976	0.994	1.481
		PCML - ML	0.843	0.847	0.028	0.968	0.990	1.457
		Jackknife						
500	8							
		PCML - ML Firth	0.948	0.945	0.018	0.936	0.965	1.175
		PCML - ML	0.895	0.891	0.017	0.908	0.952	1.175
		Jackknife	0.769	0.766	0.018	0.991	1.000	2.780
500	12							
		PCML - ML Firth	0.959	0.957	0.014	0.906	0.954	1.067
		PCML - ML	0.917	0.915	0.014	0.878	0.925	1.067
		Jackknife	0.894	0.886	0.015	0.987	0.992	2.535
1000	4							
		PCML - ML Firth	0.885	0.884	0.020	0.978	0.993	1.532
		PCML - ML	0.832	0.832	0.019	0.961	0.987	1.506
		Jackknife						
1000	8							
		PCML - ML Firth	0.953	0.951	0.013	0.928	0.973	1.155
		PCML - ML	0.900	0.898	0.012	0.854	0.929	1.156
		Jackknife	0.765	0.762	0.013	1.000	1.000	3.743
1000	12							
		PCML - ML Firth	0.959	0.958	0.010	0.891	0.948	1.059
		PCML - ML	0.917	0.916	0.010	0.815	0.887	1.058
		Jackknife	0.895	0.895	0.011	0.999	1.000	3.455

Notes: 1000 replications. PCML-ML Firth denotes the proposed estimator; PCML-ML denotes the estimator of the APE based on the PCML estimate of β and the uncorrected estimated of $\eta_i(\tilde{\gamma})$; Jackknife denotes Dhaene and Jochmans (2015)'s half-panel jackknife bias corrected estimator.

Table 3: Female labor force participation: static logit model

	Model parameters $\boldsymbol{\beta}$		Average partial effects $\boldsymbol{\mu}$		
	CML	Jackknife	CML - ML Firth	CML - ML	Jackknife
# Children 0-2	-1.183***	-1.172***	-0.096***	-0.089***	-0.106***
	(0.100)	(0.121)	(0.019)	(0.017)	(0.015)
# Children 3-5	-0.909***	-0.931***	-0.074***	-0.069***	-0.084***
	(0.111)	(0.133)	(0.021)	(0.019)	(0.016)
# Children 6-17	-0.272***	-0.271**	-0.022	-0.020	-0.024***
	(0.097)	(0.115)	(0.018)	(0.017)	(0.006)
Husband income	-0.014***	-0.011*	-0.001	-0.001	-0.001
	(0.005)	(0.006)	(0.001)	(0.001)	(0.001)
Age	1.646*	1.714	0.006	0.006	0.018
	(0.918)	(1.086)	(0.040)	(0.037)	(0.044)
Age squared	-0.224*	-0.211			
	(0.127)	(0.150)			

Notes: standard errors in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.10. CML denotes the Conditional Maximum Likelihood estimator; Jackknife denotes Hahn and Newey (2004)'s jackknife bias corrected estimator; CML-ML Firth denotes the proposed estimator; CML-ML denotes the estimator of the APE based on the CML estimate of $\boldsymbol{\beta}$ and the uncorrected estimated of $\alpha_i(\tilde{\boldsymbol{\beta}})$. Source: PSID 1979-1985.

The estimation results for the static logit model are reported in Table 3. We report both the CML and Hahn and Newey (2004)'s panel jackknife estimates of the model parameters. Despite the coefficients estimated by CML are all statistically significant at least at the 10% level, when we look at the estimated APE the model covariates do not seem to exert a significant effect on the probability of being employed, with the exception of having children between 0 and 2 years old, which reduces the probability of being employed by 9.6 percentage points, and having children between 3 and 5 years old, which reduces the probability by 7.4. It is worth noticing that CML and jackknife parameter estimates are quite similar, even though T is only 7 and the sample is around 2000. There is instead a noticeable difference in the standard errors, that are larger for the jackknife. The estimates of the APE seem to be along the same lines, whether obtained by the proposed estimator, the uncorrected CML-ML or the panel jackknife.

Table 4 reports the results for the dynamic logit specification. Here we report both the PCML and Dhaene and Jochmans (2015)'s half-panel jackknife estimators. The PCML estimator detects a strong state dependence in labor force participation of married women, as the estimated coefficient for lagged participation amounts to 1.713. In terms of APE, this is translated into an increase of 13.4 percentage points in the probability of being employed at time t for a woman who was working in $t-1$, with respect to a woman who was not working in $t-1$. The effect of the other model covariates remains null, again with the exception of the presence of young children in the family, although the APEs are now smaller. It is worth noticing that the estimated state dependence by half-panel jackknife

Table 4: Female labor force participation: dynamic logit model

	Model parameters $\boldsymbol{\theta}$		Average partial effects $\boldsymbol{\nu}$		
	PCML	Jackknife	PCML - ML Firth	PCML - ML	Jackknife
# Children 0-2	-0.909***	-1.083***	-0.060***	-0.056***	-0.107
	(0.099)	(0.141)	(0.017)	(0.016)	(0.252)
# Children 3-5	-0.555***	-0.713***	-0.037**	-0.034**	-0.073
	(0.102)	(0.149)	(0.017)	(0.016)	(0.187)
# Children 6-17	-0.173*	-0.136	-0.011	-0.011	-0.020
	(0.093)	(0.136)	(0.016)	(0.015)	(0.103)
Husband income	-0.009**	-0.006	-0.001	-0.001	-0.001
	(0.004)	(0.006)	(0.001)	(0.001)	(0.088)
Age	1.509*	-0.152	0.013	0.012	-0.022
	(0.827)	(1.269)	(0.031)	(0.029)	(0.139)
Age squared	-0.185*	-0.154			
	(0.111)	(0.172)			
Lagged participation	1.713***	2.200***	0.134***	0.127***	0.127
	(0.103)	(0.082)	(0.022)	(0.021)	(0.137)

Notes: standard errors in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.10. PCML denotes the Pseudo Conditional Maximum Likelihood estimator; Jackknife denotes Dhaene and Jochmans (2015)'s half-panel jackknife bias corrected estimator; PCML-ML Firth denotes the proposed estimator; PCML-ML denotes the estimator of the APE based on the PCML estimate of $\boldsymbol{\theta}$ and the uncorrected estimated of $\eta_i(\tilde{\boldsymbol{\theta}})$. Source: PSID 1979-1985.

is large, and that the estimated APE amounts to 12.7 percentage points, similar to that obtained with PCML - ML Firth.[11]

Our last exercise consists of a calibrated simulation study, in order to investigate the finite sample performance of the proposed estimator with a design close to a real data application.[12] For the static and dynamic models, we draw $n = 500, 1000$ women from PSID, each observed for 7 time occasions, and estimate the model parameters obtaining $\tilde{\boldsymbol{\beta}}$ by CML for the static logit model and $\tilde{\boldsymbol{\theta}}$ by PCML for the dynamic logit model. We then use $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\theta}}$ to generate data from a static or dynamic logit model, keeping the model covariates fixed and generating the error terms as a standard logistic random variables. We then re-estimate the model parameters and compute the APEs. The simulation is repplicated 1000 times.

The results for the static and dynamic logit models are reported in Tables 5 and 6, respectively. From Table 5 it emerges that the performance of the proposed estimator improves when n goes from 500 to 1000, and the results are comparable with those obtained with the simpler simulation design by Hahn and Newey (2004) in Section 4.1. Overall, the same happens for the dynamic logit as well, although in a less evident manner. Compared to the standard deviations, for both models standard errors are rather large,

---

[11]The APE, however, is not statistically significant; Dhaene and Jochmans (2015) argue that this can be expected with half-panel jackknife, which may not be very precise with short T. They recommend using their half-panel jackknife correction of the objective function instead.

[12]The structure of the calibrated simulation study is taken after Chen et al. (2018).

that however should be shrinking with larger T as suggested by the results in Section 4.

Table 5: Calibrated simulation results for the static logit model based on PSID 1979 - 1985, CML-ML Firth

	Mean ratio	Median ratio	SD	Confidence 90%	95%	SE/SD
$n = 500$						
# Children 0-2	0.917	0.917	0.019	0.804	0.892	1.212
# Children 3-5	0.941	0.941	0.018	0.916	0.956	1.233
# Children 6-17	1.220	1.208	0.017	0.749	0.863	1.225
Husband income	0.999	0.974	0.001	0.957	0.986	1.236
Age	0.853	0.877	0.003	0.973	0.989	1.322
$n = 1000$						
# Children 0-2	0.977	0.975	0.015	0.926	0.971	1.167
# Children 3-5	0.999	1.001	0.015	0.935	0.976	1.128
# Children 6-17	1.143	1.144	0.012	0.825	0.913	1.179
Husband income	1.016	1.017	0.001	0.946	0.985	1.184
Age	1.014	1.010	0.003	0.951	0.983	1.187

Notes: 1000 replications. Source: PSID 1979 - 1985.

6 Concluding remarks

So far, the literature has proposed analytical or jackknife bias corrected APE estimators. They often depend on some bias corrected estimators of the slope coefficients, which are ensured to have confidence intervals centered at their probability limit only when T grows faster than $n^{1/3}$, meaning that they attain a good finite sample performance with rather large T compared to n. This is rarely the case in microeconomic applications, where the number of subjects is often much larger than the number ot time occasions, especially in surveys with rotating sampling designs.

The method presented in this paper partly overcomes this issue by exploiting a fixed-T consistent estimator of the slope coefficients of the logit model. The proposed estimator has asymptotic bias $O(T^{-2})$, but it is shown to perform well in finite samples, even when n is much larger than T. Moreover, the bias corrected estimate of the unobserved heterogeneity based on the modified score by Firth (1993) entails a substantial improvement over the standard ML estimate with short T.

The models here presented can be estimated using the R package `cquad` and the R functions to estimate the APEs and APEs standard errors are available upon request from the Authors.

Table 6: Calibrated simulation results for the dynamic logit model based on PSID 1979 - 1985, PCML-ML Firth

		Mean ratio	Median ratio	SD	Confidence 90%	95%	SE/SD
$n = 500$							
	# Children 0-2	0.935	0.935	0.022	0.925	0.965	1.306
	# Children 3-5	0.962	0.971	0.022	0.956	0.983	1.300
	# Children 6-17	1.106	1.107	0.020	0.950	0.977	1.294
	Husband income	1.058	1.129	0.001	0.949	0.983	1.201
	Age	0.938	0.943	0.004	0.966	0.988	1.297
	Lagged participation	0.991	0.989	0.024	0.961	0.983	1.243
$n = 1000$							
	# Children 0-2	0.967	0.965	0.014	0.970	0.993	1.405
	# Children 3-5	1.022	1.028	0.014	0.984	0.998	1.454
	# Children 6-17	1.142	1.146	0.013	0.968	0.985	1.403
	Husband income	1.127	1.087	0.001	0.977	0.994	1.387
	Age	0.965	0.986	0.003	0.984	0.995	1.472
	Lagged participation	1.029	1.032	0.017	0.954	0.983	1.312

Notes: 1000 replications. Source: PSID 1979 - 1985.

References

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

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(1):102–116.

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

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.

Chen, M., Fern'andez-Val, I., and Weidner, M. (2018). Nonlinear Factor Models for Network and Panel Data. cenmap working paper CWP38/18, The Institute for Fiscal Studies.

Chernozhukov, V., Fernández-Val, I., Hahn, J., and Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2):535–580.

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

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(6):1152–1191.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

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

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

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

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

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

Stammann, A., Heiß, F., and McFadden, D. (2016). Estimating fixed effects logit models with large panel data. Technical Report G01-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft.