Conditional inference and bias reduction for partial effects estimation of fixed-effects logit models

Bartolucci, Francesco and Pigini, Claudia and Valentini, Francesco

6 October 2021

Conditional inference and bias reduction for partial effects estimation of fixed-effects logit models

Francesco Bartolucci
Università di Perugia (IT)
francesco.bartolucci@unipg.it

Claudia Pigini
Università Politecnica
delle Marche and MoFiR (IT)
c.pigini@univpm.it

Francesco Valentini
Università Politecnica
delle Marche (IT)
f.valentini@univpm.it

October 6, 2021

Abstract

We propose a multiple-step procedure to compute average partial effects (APEs) for fixed-effects panel logit models estimated by Conditional Maximum Likelihood (CML). As individual effects are eliminated by conditioning on suitable sufficient statistics, we propose evaluating the APEs at the ML estimates for the unobserved heterogeneity, along with the fixed-T consistent estimator of the slope parameters, and then reducing the induced bias in the APE by an analytical correction. The proposed estimator has bias of order $O(T^{-2})$, it performs well in finite samples and, when the dynamic logit model is considered, better than alternative plug-in strategies based on bias-corrected estimates for the slopes, especially with small n and T. We provide a real data application based on labour supply of married women.

Keywords: Average partial effects, Bias reduction, Binary panel data, Conditional Maximum Likelihood

JEL Classification: C12, C23, C25

1 Introduction

Practitioners who estimate binary choice models are often interested in quantifying the effect of some regressors 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 the number of time occasions T goes to $+\infty$ and otherwise suffers form the well-known incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000)), [1] which is why bias reduction techniques for the ML estimator of the slope parameters have been proposed in order to lower the order of the bias from $O(T^{-1})$ to $O(T^{-2})$. Related plug-in estimators of the APEs are also provided, which share the same order of bias. Analytical bias corrections 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 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.

Another popular method to overcome the incidental parameters problem is based on the conditional inference approach for fixed-effects binary logit panel data models (Andersen, 1970; Chamberlain, 1980), which admit sufficient statistics for the individual intercepts. Differently from the above-mentioned bias reduction methods, the CML method produces a fixed-T consistent estimator of the slope parameters. However, plug-in estimates of the APEs are not directly available as the parameters for the individual effects are eliminated.

We propose a multiple-step procedure to estimate the APEs in fixed-effects panel logit models that combines conditional inference approaches with a bias reduction method. The APE is evaluated at the fixed-T consistent CML estimator of the slope parameters and at the ML estimator for the unobserved heterogeneity. The bias in the APE resulting from plugging in the estimated fixed effects is then reduced from $O(T^{-1})$ to $O(T^{-2})$ by applying the analytical correction proposed by Fernández-Val (2009).

The proposed procedure cannot be directly extended to the dynamic logit model (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) for dynamic binary panel data models, which has the advantage of admitting sufficient statistics for the individual intercepts. Furthermore,

[1] We focus on large n and large T perspective, where n is the number of subjects in the sample, as APEs are often not point identified with fixed T (Chernozhukov et al., 2013).

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.

As it happens with the APE estimators based on analytical and jackknife corrections, the proposed method reduces the order of the bias from $O(T^{-1})$ to $O(T^{-2})$. Such a bias is however asymptotically negligible under rectangular array asymptotics as plug-in average-effect estimators converge at the rate $n^{-1/2}$ (Dhaene and Jochmans, 2015). Yet in spite of the asymptotic equivalence of bias-corrected and ML plug-in APE estimators, the simulation evidence provided by Dhaene and Jochmans (2015) suggests that operating some bias reduction entails a non-negligible improvement in small samples, especially with small values of T. In addition, evaluating the APE at a fixed-T consistent estimator of the regression parameters rather than at bias corrected one, gives a further advantage in finite samples. We show that the proposed combination of the conditional inference approach with bias reduction has a comparable finite sample performance to the ML and bias corrected estimators with the static logit model, while it outperforms them when the dynamic logit is considered, especially when n and T are small.

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 recall the existing bias correction strategies for APE estimators, then we illustrate the proposed methodology and its extension to accommodate the dynamic logit model; in Section 4 we investigate the finite sample performance of the proposed estimator and compare it with that of the panel jackknife and analytical bias correction strategies; in Section 5 we provide a real data application based on labour supply of married women. Finally, Section 6 concludes.

2 Average partial effects and the incidental parameters problem

We consider n units, indexed with $i = 1, \dots, n$, observed at time occasions $t = 1, \dots, T$. 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 the y_{it} are conditionally independent, given α_i and $\boldsymbol{x}_{it}$, 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 $\boldsymbol{\beta}$ is obtained by concentrating out the α_i as the solution to

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \sum_{i=1}^{n}\sum_{t=1}^{T} \log p(y_{it}|\boldsymbol{x}_{it}; \hat{\alpha}_i(\boldsymbol{\beta}), \boldsymbol{\beta}),\\ \hat{\alpha}_i(\boldsymbol{\beta}) &= \underset{\alpha_i}{\operatorname{argmax}} \sum_{t=1}^{T} \log p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}).\end{aligned}$$

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 $\underset{n\to\infty}{\operatorname{plim}} \hat{\boldsymbol{\beta}} \equiv \boldsymbol{\beta}_T \neq \boldsymbol{\beta}_0$. This is the well-known incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000). To clarify this, consider any function $f(\boldsymbol{y}_i, \boldsymbol{X}_i, \alpha_i)$ and let $\mathrm{E}_n\left[f(\boldsymbol{y}_i, \boldsymbol{X}_i, \alpha_i)\right] \equiv \underset{n\to\infty}{\operatorname{plim}} \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{y}_i, \boldsymbol{X}_i, \alpha_i)$, 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

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

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

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

where $\alpha_i(\boldsymbol{\beta})$ maximizes $\mathrm{E}_T\left[\log p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta})\right] \equiv \underset{T\to\infty}{\operatorname{plim}} \frac{1}{T}\sum_{t=1}^{T} \log p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta})$. 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 $\hat{\alpha}_i(\boldsymbol{\beta}) \xrightarrow{p} \alpha_{i0}$, with $\alpha_{i0} = \alpha_i(\boldsymbol{\beta}_0)$, and $\boldsymbol{\beta}_T \to \boldsymbol{\beta}_0$. If both $n, T \to \infty$, $\hat{\boldsymbol{\beta}}$ will be asymptotically normal. However, Hahn and Newey (2004) show that the asymptotic distribution of the ML estimator will not be centered at its probability limit if n grows faster than T.

The incidental parameters problem 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 the covariate, as

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

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 $m_{itk}(\alpha_i, \boldsymbol{\beta}, \boldsymbol{x}_{it})$ with respect to $\boldsymbol{x}_{it}$,

$$\mu_{k0} = \int m_{itk}(\alpha_{i0}, \boldsymbol{\beta}_0, \boldsymbol{x}_{it}) dG(\alpha_{i0}, \boldsymbol{x}_{it})$$

where $G(\alpha_{i0}, \boldsymbol{x}_{it})$ denotes the joint distribution of α_{i0} and $\boldsymbol{x}_{it}$. 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 small T, this estimator is plagued by two sources of bias: the first stems from the estimation error introduced by $\hat{\alpha}_i(\boldsymbol{\beta})$; 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 generated 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 $O(T^{-1})$.

The sources of bias discussed above, however, have been shown to become asymptotically negligible as, under rectangular array asymptotics, plug-in estimators of average effects converge at a rate slower than $(nT)^{-1/2}$. Dhaene and Jochmans (2015) summarize this property in their Theorem 5.1, which is based on the following rationale. Consider the infeasible estimator

$$\mu_k^* \equiv \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} m_{itk}(\alpha_{i0}, \boldsymbol{\beta}_0, \boldsymbol{x}_{it}),$$

and let μ_{ik} be the individual-specific average partial effect, with mean μ_{k0} and finite variance. Then μ_k^* can be written as

$$\mu_k^* = \frac{1}{n} \sum_{i=1}^{n} \mu_{ik} + \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T} \sum_{t=1}^{T} m_{itk}(\alpha_{i0}, \boldsymbol{\beta}_0, \boldsymbol{x}_{it}) - \mu_{ik} \right).$$

Notice that the first term converges to μ_{k0} at the rate $n^{-1/2}$, whereas the second term converges to zero at the rate $(nT)^{-1/2}$, thus implying that the infeasible APE estimator will converge no faster than the at rate $n^{-1/2}$.

From the above expression, it is straightforward to notice that any feasible average-effect estimator will converge at the same rate as μ_k^*, thus making the bias introduced by replacing α_{i0} and $\boldsymbol{\beta}_0$ with ML estimates, or their first order bias-corrected versions, asymptotically negligible. Indeed, even if it were possible to plug in fixed-T consistent estimators of α_{i0} and $\boldsymbol{\beta}_0$, the APE estimator would be asymptotically equivalent to the plug-in ML-APE estimator. However, based on their simulation evidence, Dhaene and Jochmans (2015) still suggest using some bias correction of the APE estimator in finite samples, especially when T is small. The proposed method operates such bias reduction, as well as the alternative analytical and jackknife bias corrections recalled in the following section.

3 Estimation of average partial effects

In the following, we first briefly review the existing strategies based on analytical and jackknife bias corrections, which represent the benchmark for the finite sample performance of the proposed estimator. We then illustrate the proposed methodology, which combines the consistent CML estimator of $\boldsymbol{\beta}_0$ and the analytical bias correction for the APE. Finally, we turn to the dynamic logit, for which the proposed procedure is based on a PCML estimator.

3.1 Existing strategies

The available bias reduction techniques for the estimation of APEs for fixed-effects binary choice models are based on either analytical or jackknife bias corrections.[2]

Analytical bias corrections for the APEs amount to plug-in a bias corrected estimate of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}^c = \hat{\boldsymbol{\beta}} - \hat{B}/T$, instead of the ML estimate in expression (2), along with $\hat{\alpha}_i(\hat{\boldsymbol{\beta}}^c)$. Doing so effectively removes the E component of the bias in (3), but the APE estimator is still plagued by the estimation noise in $\hat{\alpha}_i(\boldsymbol{\beta})$, giving rise to the D component. In order to remove it, the bias corrected estimator of μ_k is computed as

$$\hat{\mu}_k^c = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_{itk}(\hat{\alpha}_i(\hat{\boldsymbol{\beta}}^c), \hat{\boldsymbol{\beta}}^c, \boldsymbol{x}_{it}) - \frac{1}{nT}\sum_{i=1}^{n}\hat{D}_i, \tag{4}$$

[2]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.

where $\hat{D}_i$ is the sample counterpart of

$$D_i = \sum_{j=0}^{\infty} \mathrm{E}_T\left[\frac{\partial m_{itk}(\alpha_i, \boldsymbol{\beta})}{\partial \alpha_i}\tau_{it-j}\right] + \mathrm{E}_T\left[\frac{\partial m_{itk}(\alpha_i, \boldsymbol{\beta})}{\partial \alpha_i}\xi_i\right] + \frac{1}{2}\mathrm{E}_T\left[\frac{\partial^2 m_{itk}(\alpha_i, \boldsymbol{\beta})}{\partial \alpha_i^2}\sigma_i^2\right], \quad (5)$$

and then evaluated at $\hat{\alpha}_i(\hat{\boldsymbol{\beta}}^c)$ and $\hat{\boldsymbol{\beta}}^c$. Expressions for τ_{is}, ξ_i, and σ_i^2 for panel binary choice models are given in Fernández-Val (2009), [3] whose derivations are based on general results for static (Hahn and Newey, 2004) and dynamic (Hahn and Kuersteiner, 2011) nonlinear panel data models.[4] 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 jackknife. 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

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

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

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

[4] The expression for D_i 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).

with odd T and overlapping subpanels, as well as an alternative jackknife estimator based on the split-panel log-likelihood correction.

It is also worth pointing out that, while very general, the half- and split-panel jackknife for dynamic models proposed by Dhaene and Jochmans (2015) require the stationarity of covariates, which is rather restrictive in practice as it rules out, for instance, the use of time dummies in the model specification. On the contrary, stationarity is not required by analytical bias corrections, nor by the proposed method.

3.2 Proposed methodology

The proposed two-step strategy is based on removing the two sources of bias in (3) by (a) using the fixed-T consistent CML estimator of $\boldsymbol{\beta}$, $\tilde{\boldsymbol{\beta}}$ instead of the ML estimator $\hat{\boldsymbol{\beta}}$ and (b) reducing the order of bias of the APE plug-in estimator, induced by $\hat{\alpha}_i(\tilde{\boldsymbol{\beta}})$, from $O(T^{-1})$ to $O(T^{-2})$ by applying the analytical bias-correction of Fernández-Val (2009) reported in Equation (4).

3.2.1 Two step estimation

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

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

where the dependence of the probability on the left hand-side upon the slope parameters is suppressed to avoid abuse of notation. It is well known that the *total score* $y_{i+} = \sum_{t=1}^T y_{it}$ is a sufficient statistic for the individual intercepts α_i (Andersen, 1970; Chamberlain, 1980). The joint probability of $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})$ conditional on y_{i+} does not depend on α_i and can therefore be written as

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}) = \frac{\exp\left[\left(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp\left[\left(\sum_{t=1}^T z_t\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]}, \qquad (6)$$

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

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

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 (6). 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 since $\tilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ as $n \to \infty$.

Further we obtain estimates of the individual intercepts α_i, which are not directly available as they have been canceled out by conditioning on the total score. Our strategy is to obtain the ML estimates of α_i, for those subjects such that $0 < y_{i+} < T$, by maximizing the individual term $\sum_{t=1}^{T} \log p_{\tilde{\boldsymbol{\beta}}}(y_{it}|\boldsymbol{x}_{it}, \alpha_i)$, where $p_{\tilde{\boldsymbol{\beta}}}(y_{it}|\boldsymbol{x}_{it}, \alpha_i)$ is the logit model probability in (1) evaluated at the CML estimate $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, denoted $\hat{\alpha}_i(\tilde{\boldsymbol{\beta}})$. As well as the ML estimator, the analytical and the jackknife bias correction, our proposal leads to an APE equal to zero for the subjects whose response configurations are made of only 0s and 1s, as the marginal effects are evaluated at the ML (non-finite) estimates of α_i. This strategy is 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 $\hat{\alpha}_i(\tilde{\boldsymbol{\beta}}) \xrightarrow{p} \alpha_{i0}$ only as $T \to \infty$. [5]

In the second step, the APEs can then be obtained by simply replacing the ML estimators in (2) with $\tilde{\boldsymbol{\beta}}$ and $\hat{\alpha}(\tilde{\boldsymbol{\beta}})$ and reducing the bias from $O(T^{-1})$ to $O(T^{-2})$ by applying the bias correction proposed by Fernández-Val (2009), that is

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

where $\tilde{D}_i$ denotes the sample counterpart of (5) evaluated in $\tilde{\boldsymbol{\beta}}$ and $\hat{\alpha}(\tilde{\boldsymbol{\beta}})$. It is worth stressing that the proposed estimator exhibit the same asymptotic properties of any feasible average effect estimator under rectangular array asymptotics, as outlined in Section 2.

3.2.2 Standard errors

In order to derive an expression for the standard errors of the APEs $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_K)'$, we need to account for the variability in $\boldsymbol{x}_{it}$ and for the use of the estimated parameters $\tilde{\boldsymbol{\beta}}$

[5] It is worth mentioning that using a bias corrected estimate of α_i, such as the one proposed by Kunz et al. (2019), along with a fixed-T consistent estimator of $\boldsymbol{\beta}$ will not reduce the order of the bias of the APE estimator to $O(T^{-2})$, as it would not take care of the last component in (5). Yet Bartolucci and Pigini (2019) show that the finite sample performance of the resulting APE estimator is comparable with that of the panel jackknife, while it outperforms it with short T.

in the first step. For the latter, 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. According to Newey and McFadden (1994), 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

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{\beta},\boldsymbol{\mu}) &= \sum_{i=1}^{n}\boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{\mu}),\\ \boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{\mu}) &= \begin{pmatrix}\boldsymbol{\nabla}_{\boldsymbol{\beta}}\ell_i(\boldsymbol{\beta})\\ \boldsymbol{\nabla}_{\mu_1}g_i(\boldsymbol{\beta},\mu_1)\\ \vdots\\ \boldsymbol{\nabla}_{\mu_K}g_i(\boldsymbol{\beta},\mu_K)\end{pmatrix},\end{aligned} \tag{7}$$

and

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

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

$$\boldsymbol{W}(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})=\boldsymbol{H}(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})^{-1}\boldsymbol{S}(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})[\boldsymbol{H}(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})^{-1}]', \tag{8}$$

where

$$\boldsymbol{S}(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})=\sum_{i=1}^{n}\boldsymbol{f}_i(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})\boldsymbol{f}_i(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})'.$$

Moreover, we have that

$$\boldsymbol{H}(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}})=\sum_{i=1}^{n}\boldsymbol{H}_i(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\mu}}),$$

where

$$\boldsymbol{H}_i(\boldsymbol{\beta},\boldsymbol{\mu})=\begin{pmatrix}\boldsymbol{\nabla}_{\boldsymbol{\beta\beta}}\,\ell_i(\boldsymbol{\beta}) & \boldsymbol{O}\\ \boldsymbol{\nabla}_{\boldsymbol{\mu\beta}}\,\boldsymbol{g}_i(\boldsymbol{\beta},\boldsymbol{\mu}) & \boldsymbol{\nabla}_{\boldsymbol{\mu\mu}}\,\boldsymbol{g}_i(\boldsymbol{\beta},\boldsymbol{\mu})\end{pmatrix} \tag{9}$$

is the derivative of $\boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{\mu})$ with respect to $(\boldsymbol{\beta},\boldsymbol{\mu})$, with $\boldsymbol{O}$ denoting 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 (7) 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 (9) 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 elements equal to 2. Finally, for the computation of the block $\boldsymbol{\nabla}_{\boldsymbol{\mu}\boldsymbol{\beta}} g_i(\boldsymbol{\beta}, \boldsymbol{\mu})$ we rely on a numerical differentiation. Once the matrix in (8) 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 $\boldsymbol{W}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\mu}})$.

3.2.3 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} is

$$p(y_{it}|\boldsymbol{x}_{it}, y_{i,t-1}; \eta_i, \boldsymbol{\delta}, \gamma) = \frac{\exp\left[y_{it}(\eta_i + \boldsymbol{x}_{it}'\boldsymbol{\delta} + y_{i,t-1}\gamma)\right]}{1 + \exp(\eta_i + \boldsymbol{x}_{it}'\boldsymbol{\delta} + y_{i,t-1}\gamma)}, \tag{10}$$

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 on 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:

$$p^*(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it}'\boldsymbol{\delta} - \sum_{t=1}^T \bar{q}_{it}y_{i,t-1}\gamma + y_{i*}\gamma)}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_{t=1}^T z_t\boldsymbol{x}_{it}'\boldsymbol{\delta} - \sum_{t=1}^T \bar{q}_{it}z_{i,t-1}\gamma + z_{i*}\gamma)}, \tag{11}$$

where $y_{i*} = \sum_{t=1}^T y_{i,t-1}y_{it}$ 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 (10) 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 (10), the partial effect of covariate w_{itk} for i at time t on the probability of $y_{it} = 1$ can be written as

$$v_{itk}(\eta_i, \boldsymbol{\theta}, \boldsymbol{w}_{it}) = \begin{cases} p(y_{it} = 1|\boldsymbol{w}_{it}; \eta_i, \boldsymbol{\delta}, \gamma)\,[1 - p(y_{it} = 1|\boldsymbol{w}_{it}; \eta_i, \boldsymbol{\delta}, \gamma)]\,\delta_k, & w_{itk} \text{ continuous,} \\ p(y_{it} = 1|\boldsymbol{w}_{it,-k}, w_{itk} = 1; \eta_i, \boldsymbol{\delta}, \gamma) - \\ \qquad p(y_{it} = 1|\boldsymbol{w}_{it,-k}, w_{itk} = 0; \eta_i, \boldsymbol{\delta}, \gamma), & w_{itk} \text{ discrete,} \end{cases}$$

where $\boldsymbol{w}_{it,-k}$ again denotes the the vector $\boldsymbol{w}_{it}$ excluding w_{itk}, and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \gamma)'$. This expression may also be used to compute the APE of the lagged response variable. Notice that this function does not depend on $\bar{\boldsymbol{\delta}}$, since the probability in (10) 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

$$\nu_{k0} = \int v_{itk}(\eta_{i0}, \boldsymbol{\theta}_0, \boldsymbol{w}_{it})dG(\eta_{i0}, \boldsymbol{w}_{it}).$$

where $dG(\eta_{i0}, \boldsymbol{w}_{it})$ denotes the joint distribution of η_{i0} and w_{it}.

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 first step in Section 3.2.1. Here, however, the CML estimation of $\boldsymbol{\theta}$ based on (11) 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

$$\ell(\bar{\boldsymbol{\delta}}) = \sum_{i=1}^n \mathbf{I}(0 < y_{i+} < T)\ell_i(\bar{\boldsymbol{\delta}}),$$

where

$$\ell_i = \log \frac{\exp\left[\left(\sum_{t=1}^T y_{it}\boldsymbol{x}_{it}\right)'\bar{\boldsymbol{\delta}}\right]}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp\left[\left(\sum_{t=1}^T z_t\boldsymbol{x}_{it}\right)'\bar{\boldsymbol{\delta}}\right]},$$

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

$$\ell_i(\bar{\eta}_i) = \sum_{t=1}^T \log \frac{\exp\left[y_{it}(\bar{\eta}_i + \boldsymbol{x}_{it}'\check{\boldsymbol{\delta}})\right]}{1+\exp(\bar{\eta}_i + \boldsymbol{x}_{it}'\check{\boldsymbol{\delta}})},$$

where $\check{\boldsymbol{\delta}}$ is fixed. The probability $\bar{q}_{it}$ in (11) 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 conditional log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{I}(0 < y_{i+} < T) \log p^*_{\check{\boldsymbol{q}}_i}(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}),$$

where $p^*_{\check{\boldsymbol{q}}_i}(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+})$ is the joint probability in (11) 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.[6] 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. Given the estimator $\tilde{\boldsymbol{\theta}}$, we recover ML estimates $\hat{\eta}_i(\tilde{\boldsymbol{\theta}})$ maximizing the individual log-likelihood based on Equation (10).

Finally, in step three, the APEs can then be estimated by plugging $\tilde{\eta}_i(\tilde{\boldsymbol{\theta}})$ and $\tilde{\boldsymbol{\theta}}$ in the APE formulation and applying the same correction shown in Section 3.2.1, so as to obtain

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

Standard errors for $\tilde{\nu}_k$ can be obtained exactly in the same way as illustrated in Section 3.2.2 with the appropriate change of notation.

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

4 Monte Carlo 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 $\alpha_i \sim N(0, 1)$, the error terms 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 $n = 100, 500$, $T = 4, 8, 12$. The coefficient β is equal to 1 across all the scenarios and the number of replications is 1000. Fernández-Val (2009) considers the same scenarios with only $n = 100$ and Hahn and Newey (2004) consider only $n = 100$ and $T = 4, 8$.

Table 1 reports the simulation results for the APE estimators related to the regressor x in each scenario. We compare the finite sample performance of the proposed APE estimator (denoted by CML-BC) with: (a) the ML plug-in estimator (ML); (b) Hahn and Newey (2004)'s jackknife bias corrected estimator (Jackknife-BC); (c) the ML estimator with the analytical bias correction (Analytical-BC) provided by Fernández-Val (2009), also mentioned in the previous section. For each scenario, we report the mean and the median of the ratio $\tilde{\mu}/\mu^*$, the standard deviation of $\tilde{\mu}$, the rejection frequency at 5% and 10% nominal value, and the mean ratio between the estimator standard error and standard deviation.[7]

From Table 1, it emerges that the proposed estimator (CML-BC) has a good finite sample performance with both small n and T. 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. The proposed procedure, the Jackknife-BC, and the Analytical-BC estimator exhibit a similar behavior in all the scenarios, which is also in line with the results reported by Hahn and Newey (2004) for the probit model. All

[7] ML standard errors are computed for Hahn and Newey (2004)'s Jackknife-BC estimator.

in all, though, it emerges that the asymptotic equivalence of the plug-in bias-corrected estimators considered with the ML APE estimator is reflected by their small sample performance as well with this particular design.

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

n	T		Mean ratio	Median ratio	SD	Rejection rate 5%	Rejection rate 10%	SE/SD
100	4							
		ML	0.988	0.993	0.072	0.048	0.102	1.006
		Jackknife - BC	0.988	0.990	0.077	0.050	0.090	1.001
		Analytical - BC	0.926	0.932	0.067	0.071	0.132	0.938
		CML - BC	0.929	0.934	0.068	0.025	0.071	1.126
100	8							
		ML	0.998	0.999	0.028	0.046	0.101	0.985
		Jackknife - BC	0.999	1.000	0.028	0.036	0.090	1.026
		Analytical - BC	0.984	0.985	0.028	0.046	0.100	0.973
		CML - BC	0.985	0.986	0.028	0.039	0.086	1.025
100	12							
		ML	0.992	0.996	0.016	0.066	0.127	0.906
		Jackknife - BC	0.996	0.999	0.016	0.056	0.110	0.936
		Analytical - BC	0.988	0.992	0.016	0.045	0.106	0.965
		CML - BC	0.989	0.993	0.016	0.042	0.101	0.986
500	4							
		ML	1.004	0.997	0.032	0.056	0.094	1.004
		Jackknife - BC	1.010	1.004	0.035	0.056	0.113	0.994
		Analytical - BC	0.940	0.934	0.030	0.091	0.154	0.931
		CML - BC	0.943	0.936	0.031	0.030	0.087	1.120
500	8							
		ML	0.996	0.996	0.013	0.066	0.122	0.929
		Jackknife - BC	0.997	0.997	0.014	0.059	0.102	0.960
		Analytical - BC	0.982	0.982	0.013	0.080	0.139	0.909
		CML - BC	0.983	0.983	0.013	0.059	0.124	0.966
500	12							
		ML	0.999	1.000	0.007	0.075	0.126	0.910
		Jackknife - BC	1.003	1.002	0.007	0.066	0.113	0.946
		Analytical - BC	0.995	0.996	0.007	0.053	0.114	0.964
		CML - BC	0.996	0.996	0.007	0.042	0.104	0.990

4.2 Dynamic logit

For the dynamic logit model, the simulation design is similar to that by Honoré and Kyriazidou (2000). The data generating process, for $i = 1, \ldots, n$, is as follows

$$\begin{aligned} y_{it} &= \mathbf{I}(\eta_i + y_{i,t-1}\gamma + x_{it}\delta + \upsilon_{it} > 0), \quad t = 1, \ldots, T, \\ y_{i0} &= \mathbf{I}(\eta_i + x_{i0}\delta + \upsilon_{i0} > 0), \end{aligned}$$

where $x_{it} \sim N(0, \pi^2/3)$ and υ_{it} follows a standard logistic distribution, for $t = 0, \ldots, T$. The individual heterogeneity is generated as $\eta_i = \sum_{t=0}^{3} x_{it}/4$. We consider the same scenarios as for the static logit model, that are $n = 100, 500$, $T = 4, 8, 12$. The coefficient

γ is equal to 0.5 and δ is equal to 1 across all the scenarios and the number of replications is 1000.

Tables 2 and 3 report the simulation results for the APE estimators related to variations in x, denoted as ν_x, and in y_{t-1}, denoted ν_y, respectively. We compare the finite sample performance of the proposed APE estimator (PCML-BC) with: (a) the ML plug-in estimator; (b) Dhaene and Jochmans (2015)'s half-panel jackknife bias corrected estimator (Jackknife-BC); (c) the analytically bias-corrected estimator (Analytical-BC) by Fernández-Val (2009). It must be noted that the half-panel Jackknife-BC estimator cannot be computed for $T = 4$. Again we report the mean and the median of the ratio $\tilde{\nu}/\nu^*$, the standard deviation of $\tilde{\nu}$, the rejection frequency at 5% and 10% nominal value, and the mean ratio between the estimator standard error and standard deviation.[8]

From Table 2 it emerges that the proposed estimator outperforms the uncorrected ML and Analytical-BC when T is equal to 4, whereas they have comparable performance with larger values of T, also considering the Jackknife-BC. Furthermore, the proposed methodology seems to provide the most reliable confidence intervals among the examined estimators. In this regard, it is worth noticing that when $T = 4$, all the estimators provide poor coverage. By contrast, the proposed procedure offers a remarkable advantage over the ML, Analytical-BC, and Jackknife-BC when it comes to the estimation of the APE related to the state dependence parameter in almost all the scenarios considered (see Table 3), in terms of mean ratio and rejection frequencies.

5 Empirical application

We apply our proposed formulation to the problem of estimating the labour 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 $T = 6$ time occasions, from 1980 to 1985, further to an additional observation in 1979 exploited as initial condition in dynamic models. 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, and on the husband's income. We also specify a dynamic logit model, that is we include lagged participation in the set of model covariates.

The estimation results for the static logit model are reported in Table 4. We re-

[8]ML standard errors are computed for Dhaene and Jochmans (2015)'s half-panel Jackknife-BC estimator. They are different from those proposed bu the authors, who suggested using the cross-sectional variance of the within-group average effects, which, however, had a worse performance in finite samples than the ML ones.

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

n	T		Mean ratio	Median ratio	SD	Rejection Rate 5%	Rejection Rate 10%	SE/SD
100	4							
		ML	0.917	0.917	0.014	0.261	0.340	0.774
		Analytical - BC	0.620	0.690	0.038	0.978	0.989	0.265
		PCML - BC	0.889	0.886	0.013	0.145	0.254	1.117
100	8							
		ML	0.988	0.989	0.008	0.065	0.116	0.891
		Jackknife - BC	1.056	1.057	0.011	0.165	0.248	0.862
		Analytical - BC	0.980	0.981	0.008	0.074	0.128	0.899
		PCML - BC	0.977	0.977	0.008	0.069	0.129	0.939
100	12							
		ML	0.998	0.997	0.006	0.057	0.117	0.844
		Jackknife - BC	1.013	1.012	0.007	0.081	0.143	0.855
		Analytical - BC	0.998	0.998	0.006	0.029	0.075	0.933
		PCML - BC	0.991	0.991	0.006	0.044	0.099	0.913
500	4							
		ML	0.912	0.913	0.006	0.691	0.780	0.819
		Analytical - BC	0.706	0.713	0.007	1.000	1.000	0.577
		PCML - BC	0.887	0.886	0.005	0.741	0.843	1.185
500	8							
		ML	0.990	0.990	0.004	0.081	0.147	0.851
		Jackknife - BC	1.060	1.061	0.005	0.545	0.650	0.803
		Analytical - BC	0.982	0.982	0.004	0.131	0.206	0.860
		PCML - BC	0.979	0.979	0.004	0.142	0.229	0.895
500	12							
		ML	0.996	0.996	0.003	0.046	0.089	0.894
		Jackknife - BC	1.013	1.013	0.003	0.117	0.201	0.868
		Analytical - BC	0.996	0.996	0.003	0.025	0.067	0.984
		PCML - BC	0.990	0.989	0.003	0.055	0.106	0.970

Table 3: Simulation results for $\tilde{\nu}_y$, dynamic logit model

n	T		Mean ratio	Median ratio	SD	Confidence 90%	Confidence 95%	SE/SD
100	4							
		ML	-1.239	-1.245	0.042	0.862	0.972	1.611
		Analytical - BC	1.166	1.140	0.053	0.101	0.169	0.714
		PCML - BC	0.849	0.839	0.055	0.025	0.063	1.195
100	8							
		ML	-0.207	-0.203	0.029	0.883	0.932	1.050
		Jackknife - BC	0.790	0.773	0.039	0.320	0.407	0.531
		Analytical - BC	0.896	0.899	0.031	0.092	0.169	0.877
		PCML - BC	0.987	0.994	0.034	0.062	0.115	0.968
100	12							
		ML	0.197	0.197	0.024	0.790	0.866	0.904
		Jackknife - BC	0.933	0.928	0.029	0.228	0.306	0.616
		Analytical - BC	0.955	0.952	0.025	0.077	0.131	0.912
		PCML - BC	1.004	0.998	0.027	0.075	0.128	0.933
500	4							
		ML	-1.231	-1.222	0.019	1.000	1.000	1.585
		Analytical - BC	1.026	1.025	0.019	0.086	0.158	0.866
		PCML - BC	0.873	0.872	0.025	0.027	0.067	1.198
500	8							
		ML	-0.203	-0.203	0.013	1.000	1.000	1.062
		Jackknife - BC	0.796	0.791	0.017	0.463	0.546	0.546
		Analytical - BC	0.905	0.903	0.014	0.109	0.180	0.899
		PCML - BC	0.993	1.000	0.015	0.057	0.104	0.996
500	12							
		ML	0.185	0.186	0.011	1.000	1.000	0.929
		Jackknife - BC	0.929	0.925	0.012	0.255	0.339	0.636
		Analytical - BC	0.945	0.949	0.011	0.087	0.156	0.931
		PCML - BC	0.991	0.995	0.012	0.062	0.110	0.962

Table 4: Female labour force participation: static logit model

Labour Force Participation	Model parameters $\boldsymbol{\beta}$			
	ML	Jackknife BC	Analytical BC	CML
# Children 0-2	-1.331***	-1.051***	-1.090***	-1.092***
	(0.145)	(0.145)	(0.117)	(0.109)
# Children 3-5	-0.922***	-0.706***	-0.755***	-0.756***
	(0.147)	(0.147)	(0.112)	(0.103)
# Children 6-17	-0.193	-0.141	-0.157*	-0.157*
	(0.123)	(0.123)	(0.088)	(0.081)
Husband income	-0.011*	-0.005	-0.009**	-0.009***
	(0.006)	(0.006)	(0.004)	(0.004)
	Average partial effects $\boldsymbol{\mu}$			
	ML	Jackknife BC	Analytical BC	CML-BC
# Children 0-2	-0.091***	-0.092***	-0.088***	-0.089***
	(0.009)	(0.009)	(0.008)	(0.021)
# Children 3-5	-0.063***	-0.066***	-0.061***	-0.061***
	(0.009)	(0.010)	(0.008)	(0.022)
# Children 6-17	-0.013	-0.012	-0.013**	-0.013
	(0.009)	(0.009)	(0.006)	(0.019)
Husband income	-0.001	-0.001	-0.001**	-0.001
	(0.001)	(0.001)	(0.000)	(0.001)

port the ML, Hahn and Newey (2004)'s panel Jackknife-BC, Hahn and Newey (2004)'s Analytical-BC and CML estimates of the model parameters. The CML, Analytical-BC, and Jackknife-BC estimates of the parameters are all similar to each other and smaller (in absolute value) than the uncorrected ML ones; they suggest a negative effect on labour participation of having children younger than 17 in the household as well as of the level of the husband's income. The estimated APEs obtained with the proposed method suggest that having an additional child between 0 and 2 reduces the probability of working by 8.9 percentage points, and having a child between 3 and 5 years old reduces the employment probability by 6.1 percentage points. The APE estimates obtained with the Analytical-BC and Jackknife-BC estimators point toward the same results, with the exception of having children between 6 and 17 years old, which appear to be not statistically significant, according to our procedure.

Table 5 reports the results for the dynamic logit specification. Here we report the ML, Dhaene and Jochmans (2015)'s half-panel Jackknife-BC, Fernández-Val (2009)'s Analytical-BC and PCML estimates of the model parameters. The effect of the exogenous model covariates is now smaller and all the APE estimates suggest a negative and statistically significant effect of having children between 0 and 5 years old in the household.

The PCML estimator detects a strong state dependence in labour force participation of married women, as the estimated coefficient for lagged participation amounts to 1.706.

Table 5: Female labour force participation: dynamic logit model

Labour force participation	Model parameters $\boldsymbol{\theta}$			
	ML	Jackknife BC	Analytical BC	PCML
# Children 0-2	-1.269***	-0.895***	-0.930***	-0.912***
	(0.141)	(0.141)	(0.122)	(0.095)
# Children 3-5	-0.823***	-0.503***	-0.532***	-0.503***
	(0.141)	(0.141)	(0.117)	(0.091)
# Children 6-17	-0.173	-0.019	-0.106	-0.092
	(0.117)	(0.117)	(0.092)	(0.074)
Husband income	-0.011*	-0.005	-0.009**	-0.008**
	(0.006)	(0.006)	(0.004)	(0.004)
Lagged Participation	0.569***	2.107***	1.319***	1.706***
	(0.081)	(0.081)	(0.082)	(0.103)
	Average partial effects $\boldsymbol{\nu}$			
	ML	Jackknife BC	Analytical BC	PCML-BC
# Children 0-2	-0.086***	-0.097***	-0.075***	-0.064***
	(0.009)	(0.008)	(0.008)	(0.016)
# Children 3-5	-0.056***	-0.059***	-0.043***	-0.035**
	(0.009)	(0.008)	(0.008)	(0.015)
# Children 6-17	-0.012	-0.009	-0.009	-0.006
	(0.008)	(0.007)	(0.006)	(0.012)
Husband income	-0.001	-0.001	-0.001**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Lagged Participation	0.041***	0.124***	0.121***	0.152***
	(0.008)	(0.030)	(0.006)	(0.021)

In terms of APE, this is translated into an increase of 15.2 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 $t-1$.

6 Conclusion

We develop a multiple-step procedure to compute APEs for fixed-effects logit models that are estimated by CML. Our strategy amounts to building a plug-in APE estimator based on the fixed-T consistent CML estimator of the slope parameters and bias corrected estimates of APEs.

The proposed estimator is asymptotically equivalent to the plug-in ML and alternative bias-corrected APE estimators, and it exhibits a comparable finite sample performance when the static logit model is considered. On the contrary, the proposed approach for the dynamic logit model has a remarkable advantage in finite samples with small T. In this respect, the multiple-step procedure here developed could be particularly useful to practitioners who often deal with short-T datasets, such as rotated surveys, and/or highly unbalanced panels.

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*, 170:102–116.

Bartolucci, F. and Pigini, C. (2019). Partial effects estimation for fixed-effects logit panel data models. MPRA Paper 92243, University Library of Munich, Germany.

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.

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

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*, 82:991–1030.

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

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

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

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.

Kunz, J., Staub, K. E., and Winkelmann, R. (2019). Predicting fixed effects in panel probit models. *Monash Business School*, 19(10).

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

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

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

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.