UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

PARTIAL EFFECTS ESTIMATION FOR FIXED-EFFECTS LOGIT PANEL DATA MODELS

FRANCESCO BARTOLUCCI CLAUDIA PIGINI

QUADERNO DI RICERCA n. 431
ISSN: 2279-9575

Agosto 2018

Abstract

We develop a multiple-step procedure for the estimation of point and average partial effects in fixed-effects logit panel data models that admit sufficient statistics for the incidental parameters. In these models, estimates of the individual effects are not directly available and have to be recovered by means of an additional step. We also derive a standard error formulation for the average partial effects. We study the finite-sample properties of the proposed estimator by simulation and provide an application based on unionised workers.

JEL Class.: C12, C23, C25
Keywords: Keywords: Partial effects, Logit model, Quadratic Exponential model, Conditional Maximum Likelihood

Partial effects estimation for fixed-effects logit panel data models

Francesco Bartolucci *Claudia Pigini*

1 Introduction

The fixed-effects approach is particularly attractive in modelling panel data, as it does not require distributional assumptions for the individual unobserved heterogeneity, which is also allowed to be correlated with the model covariates in a nonparametric way. In nonlinear models, however, the inclusion of individual intercepts among the model covariates typically gives rise to the well-known incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000), which makes the Maximum Likelihood (ML) estimator inconsistent for all the model parameters if T is fixed.

For the standard fixed-effects logit model, the incidental parameters problem can be solved by conditioning on simple sufficient statistics for the individual intercepts (Andersen, 1970; Chamberlain, 1980) and the model parameters can be consistently estimated by Conditional Maximum Likelihood (CML). If, however, a dynamic logit is specified (Hsiao, 2005), namely the lagged dependent variable is included among the model covariates, CML inference is not viable in a simple form.[1] 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.

The CML and pseudo-CML estimators of the static and dynamic logit models have been shown to perform quite well in finite samples, even with very short T. However, one drawback that may have limited their application in practice is the lack of a back-of-the-envelope calculation of Partial Effects (PE), as estimates of the individual effects are not directly available with CML. Here we develop a multiple-step procedure for the computation of point PE and Average PE (APE) for the static and dynamic logit models. Structural parameters are estimated by CML in the first steps and used in a second step to estimate the individual effects by ML. While the ML estimator is not consistent for the individual parameters with fixed T, we show by simulation that it provides a very close approximation of the true APE. We derive the formulation for the standard errors of the APE, which does not depend on the estimates of the individual intercepts and therefore holds with fixed T.

[1] 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 $\sqrt{n}$.

The remainder of the paper is organised as follows: in Section 2 we illustrate the multiple-step procedure for the static and dynamic logit model and derive the APE standard error formulation. In Section 3 we investigate the finite-sample performance of the proposed estimator by simulation and in Section 4 we provide a real data application based on unionised workers. Finally, Section 5 concludes.

2 Partial effects estimation

We consider n units, indexed with $i = 1, \dots, n$, observed at time occasions $t = 1, \dots, T$.[2] Let y_{it} be the binary response variable for unit i at occasion t and $\boldsymbol{x}_{it}$ the corresponding vector of K covariates. In the following, we first consider the static logit model and then the dynamic logit, for which the estimation of partial effects involves an additional step.

2.1 Static logit model

Consider the logit formulation

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

where α_i is the individual specific intercept, bx_{it} is vector of strictly exogenous covariates, and $\boldsymbol{\beta}$ collects the regression parameters. 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 $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})$ conditional on y_{i+} does not depend on α_i and can therefore be written as

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}) = \frac{\exp\left[\left(\sum_t y_{it}\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp\left[\left(\sum_t z_t\boldsymbol{x}_{it}\right)'\boldsymbol{\beta}\right]}, \tag{2}$$

where the individual intercepts α_i have been cancelled out (see also Bartolucci and Pigini, 2017, for details).

Considering expression (1), the PE 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

$$f_{itk}(\alpha_i, \boldsymbol{\beta}, \boldsymbol{x}_{it}) = \begin{cases} p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it})\left[1 - p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it})\right]\beta_k, & x_{itk} \text{ continuous} \\ p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it,-k}, x_{itk} = 1) - \\ \qquad p(y_{it} = 1|\alpha_i, \boldsymbol{x}_{it,-k}, x_{itk} = 0), & x_{itk} \text{ discrete} \end{cases} \tag{3}$$

where $\boldsymbol{x}_{it,-k}$ denotes the subvector of all covariates but x_{itk}. The 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 $\boldsymbol{x}_{it}$:

$$\phi_k = \mathrm{E}_{\boldsymbol{x}_{it}}\left[f_{itk}(\alpha_i, \boldsymbol{\beta}, \boldsymbol{x}_{it})\right]. \tag{4}$$

[2] In order to keep the notation clearer only balanced panel data are considered, however the developed methods apply to unbalanced panel data as well with obvious adjustments.

Estimation of the PEs in (3) and of the APEs in (4) requires an estimate of the individual intercepts, which is not directly available if the parameters of the logit model are estimated by CML based on (2). We therefore propose the following two-step procedure:

Step 1. We estimate the structural parameters of the logit model in (1) by CML. The log-likelihood function is

$$\ell(\boldsymbol{\beta}) = \sum_i \mathbf{I}(0 < y_{i+} < T) \log p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}), \tag{5}$$

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 (2). The above function can be maximised 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 $\hat{\boldsymbol{\beta}}$.

Step 2. We obtain the ML estimates of the individual intercepts α_i, for those subjects such that $0 < y_{i+} < T$, by maximising the individual log-likelihood functions

$$\ell_i(\alpha_i|\hat{\boldsymbol{\beta}}) = \sum_t \log p_{\hat{\boldsymbol{\beta}}}(y_{it}|\alpha_i, \boldsymbol{x}_{it}),$$

where $p_{\hat{\boldsymbol{\beta}}}(y_{it}|\alpha_i, \boldsymbol{x}_{it})$ is the logit model probability in (1) evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. The resulting estimates of the individual intercepts $\hat{\alpha}_i$ will depend on $\hat{\boldsymbol{\beta}}$, which we write as $\hat{\alpha}_i(\hat{\boldsymbol{\beta}})$. The PEs and the APEs can then be estimated by simply substituting $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\alpha_i = \hat{\alpha}_i(\hat{\boldsymbol{\beta}})$ in (3) and by taking the sample average

$$\hat{\phi}_k = \frac{1}{nT} \sum_i \sum_t f_{itk}(\hat{\alpha}_i(\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}, \boldsymbol{x}_{it}) \tag{6}$$

to estimate ϕ_k in (4).

2.2 Dynamic logit

For the dynamic logit model (Hsiao, 2005), the conditional probability of y_{it} being equal to 1 is

$$p(y_{it}|\eta_i, \boldsymbol{x}_{it}, y_{i,t-1}) = \frac{\exp\left[y_{it}(\eta_i + \boldsymbol{x}_{it}'\boldsymbol{\delta} + y_{i,t-1}\gamma)\right]}{1 + \exp(\eta_i + \boldsymbol{x}_{it}'\boldsymbol{\delta} + y_{i,t-1}\gamma)}, \tag{7}$$

where γ is the regression coefficient for the lagged response variable that measures the true state dependence. The dynamic logit model does not admit sufficient statistics for the incidental parameters and Bartolucci and Nigro (2012) propose a pseudo-CML estimator based on approximating the dynamic logit model by a QE model (Cox, 1972). 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:

$$p^*(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp(\sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\delta} - \sum_t \bar{q}_{it}y_{i,t-1}\gamma + y_{i*}\gamma)}{\sum\limits_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\delta} - \sum_t \bar{q}_{it}z_{i,t-1}\gamma + z_{i*}\gamma)}, \tag{8}$$

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-series Taylor approximation of the log-likelihood based on (7) 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 $\bar{q}_{it} = \exp(\bar{\eta}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\delta}})/\left[1 + \exp(\bar{\eta}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\delta}})\right]$

Expressions for PEs 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 (7), the PE of covariate w_{itk} for i at time t on the probability of $y_{it} = 1$ can be written as

$$f_{itk}(\eta_i, \boldsymbol{\theta}, \boldsymbol{w}_{it}) = \begin{cases} p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it})\left[1 - p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it})\right]\delta_k, & w_{itk} \text{ continuous} \\ p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it,-k}, w_{itk} = 1) - \\ \qquad p(y_{it} = 1|\eta_i, \boldsymbol{w}_{it,-k}, w_{itk} = 0), & w_{itk} \text{ discrete} \end{cases} \tag{9}$$

where $\boldsymbol{w}_{it,-k}$ again denotes the the vector $\boldsymbol{w}_{it}$ excluding w_{itk}, and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \gamma)'$. Notice that the PE function does not depend on $\bar{\boldsymbol{\delta}}$, since the probability in (7) does not depend on $\bar{q}_{it}$. The APE of the k-th covariate can then be obtained by taking the expected value with respect to $\boldsymbol{w}_{it}$ of $f_{itk}(\eta_i, \boldsymbol{\theta}, \boldsymbol{w}_{it})$:

$$\psi_k = \mathrm{E}_{\boldsymbol{w}_{it}}\left[f_{itk}(\eta_i, \boldsymbol{\theta}, \boldsymbol{w}_{it})\right]. \tag{10}$$

As for the static logit model, the estimation of PEs in (9) and APEs in (10) requires an estimate of η_i, which we recover in a similar manner as in step 2 in Section 2.1. Here, however, the CML estimation of $\boldsymbol{\theta}$ based on (8) relies on a preliminary step in order to obtain $\bar{q}_{it}$ and the estimation of PEs and APEs is thus based on the following three-step procedure:

Step 1 A preliminary estimate of $\bar{\boldsymbol{\delta}}$ is obtained by maximising the conditional log-likelihood

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

where

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

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

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

where $\tilde{\boldsymbol{\delta}}$ is fixed. The probability $\bar{q}_{it}$ in (8) can the be estimated by

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

Step 2 We estimate $\boldsymbol{\theta}$ by maximising the following conditional log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_i \mathbf{I}(0 < y_{i+} < T) \log p^*_{\tilde{\boldsymbol{q}}_i}(\boldsymbol{y}_i | \boldsymbol{X}_i, y_{i0}, y_{i+}), \tag{11}$$

where $p^*_{\tilde{\boldsymbol{q}}_i}(\boldsymbol{y}_i | \boldsymbol{X}_i, y_{i0}, y_{i+})$ is the joint probability in (8) evaluated at $\tilde{\boldsymbol{q}}_i = (\tilde{q}_{i1}, \ldots, \tilde{q}_{iT})'$. The above function can be easily maximised with respect to $\boldsymbol{\theta}$ by the Newton-Raphson algorithm, so as to obtain the pseudo-CML estimator $\hat{\boldsymbol{\theta}}$.

Step 3 We obtain the ML estimates of the individual intercepts η_i, for those subjects such that $0 < y_{i+} < T$, by maximising the following log-likelihood function

$$\ell_i(\eta_i | \hat{\boldsymbol{\theta}}) = \log p_{\hat{\boldsymbol{\theta}}}(y_{it} | \eta_i, \boldsymbol{x}_{it}, y_{it-1}),$$

where $p_{\hat{\boldsymbol{\theta}}}(y_{it} | \eta_i, \boldsymbol{x}_{it}, y_{it-1})$ is the dynamic logit model probability in (7) evaluated at $\hat{\boldsymbol{\theta}}$. The resulting estimates of the individual intercepts $\hat{\eta}_i$ will depend on $\hat{\boldsymbol{\theta}}$, $\hat{\eta}_i(\hat{\boldsymbol{\theta}})$. The PEs and the APEs can then be estimated by simply substituting $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\eta_i = \hat{\eta}_i(\hat{\boldsymbol{\theta}})$ in (9) and taking the sample average to estimate ψ_k in (10).

2.3 Standard errors

In order to derive an expression for the standard errors of the APEs $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_K)'$ and $\hat{\boldsymbol{\psi}} = (\hat{\psi}_1, \ldots, \hat{\psi}_K)'$, we need to account for the use of the estimated parameters $\hat{\boldsymbol{\beta}}$ in step 1 and $\hat{\boldsymbol{\theta}}$ in step 2 for the static and dynamic logit model, respectively. In the following we derive the standard error formulation for the first model and the same strategy can be directly applied to the dynamic logit as well, since the preliminary estimates obtained in step 1 for the dynamic logit model do not enter the probability in (7), used for the formulation of PEs in (9) and APEs in (10)

We rely on the Generalised Method of Moments (GMM) approach by Hansen (1982) and also implemented by Bartolucci and Nigro (2012) for the QE model. The proposed method consists in presenting the proposed multi-step procedure as the solution of the system of estimating equations

$$\boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \mathbf{0},$$

where

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

$$\boldsymbol{g}_i(\boldsymbol{\beta}, \boldsymbol{\phi}) = \begin{pmatrix} \boldsymbol{\nabla}_{\boldsymbol{\beta}} \ell_i(\boldsymbol{\beta}) \\ \boldsymbol{\nabla}_{\phi_1} m_i(\boldsymbol{\beta}, \phi_1) \\ \vdots \\ \boldsymbol{\nabla}_{\phi_K} m_i(\boldsymbol{\beta}, \phi_K) \end{pmatrix}, \tag{12}$$

and

$$m_i(\boldsymbol{\beta}, \phi_k) = \frac{1}{T} \sum_t \left[f_{itk}(\boldsymbol{\beta}, \alpha_i(\boldsymbol{\beta})) - \phi_k \right]^2, \ k = 1, \ldots, K.$$

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

$$\boldsymbol{W}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = \boldsymbol{H}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})^{-1}\boldsymbol{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})[\boldsymbol{H}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})^{-1}]', \tag{13}$$

where

$$\boldsymbol{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = \sum_i \mathbf{I}(0 < y_{i+} < T)\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})',$$

$$\boldsymbol{H}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = \sum_i \mathbf{I}(0 < y_{i+} < T)\boldsymbol{H}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}),$$

and

$$\boldsymbol{H}_i(\boldsymbol{\beta}, \boldsymbol{\phi}) = \begin{pmatrix} \boldsymbol{\nabla}_{\boldsymbol{\beta}\boldsymbol{\beta}}\ell_i(\boldsymbol{\beta}) & \boldsymbol{O} \\ \boldsymbol{\nabla}_{\boldsymbol{\phi}\boldsymbol{\beta}}\boldsymbol{m}_i(\boldsymbol{\beta}, \boldsymbol{\phi}) & \boldsymbol{\nabla}_{\boldsymbol{\phi}\boldsymbol{\phi}}\boldsymbol{m}_i(\boldsymbol{\beta}, \boldsymbol{\phi}) \end{pmatrix}, \tag{14}$$

is the derivative of $\boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{\phi})$ with respect to $(\boldsymbol{\beta}, \boldsymbol{\phi})$, where $\boldsymbol{O}$ denotes a $K \times K$ matrix of zeros and $\boldsymbol{m}_i(\boldsymbol{\beta}, \boldsymbol{\phi})$ collects $m_i(\boldsymbol{\beta}, \phi_k)$, for $k = 1, \ldots, K$ Expressions for the derivatives in $\boldsymbol{g}_i(\boldsymbol{\beta}, \boldsymbol{\phi})$ and $\boldsymbol{H}_i(\boldsymbol{\beta}, \boldsymbol{\phi})$ are given in Appendix. Once the matrix in (13) is computed, the standard errors for the APEs $\hat{\boldsymbol{\phi}}$ may be obtained by taking the square root of the elements in the main diagonal of the lower right submatrix of $\boldsymbol{W}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$.

3 Simulation study

We evaluate the finite sample performance of the estimators for the APEs in (4) and (10) and of the procedure to obtain the standard errors by means of a simulation study, where data are generated as follows:

$$y_{it} = \mathbf{I}(\alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \gamma y_{i,t-1}\varepsilon_{it} > 0), \quad i = 1, \ldots, n, \; t = 2, \ldots, T,$$

with initial condition

$$y_{i1} = \mathbf{I}(\alpha_i + x_{i11}\beta_1 + x_{i12}\beta_2 + \varepsilon_{i1} > 0).$$

In our benchmark design x_{it1} is a standard normal random variable, and x_{it2} is a binary variable generated as $x_{it2} = \mathbf{I}(x^*_{it2} > 0)$, with x^*_{it2} following a standard normal distribution, for $t = 1, \ldots, T$. Furthermore, the individual effects are generated as $\alpha_i = 1/4\sum_{t=1}^{4} x_{it1}$ and the terms ε_{it} follow a standard logistic distribution. In a second design, we let $x_{it1} \sim \chi^2_1$ and $x^*_{it2} \sim \chi^2_1$, centred at -0.1 so that the sample average of x_{it2} is about 0.765.

For each design, the study is based on four scenarios. For the static logit model, we hold $\gamma = 0$ and set (β_1, β_2) first to $(0, 0)$ and then to $(1, -1)$. For the dynamic logit model we set the regression parameters to $(1, -1)$ while γ takes values 0 and 1. For each scenario, we ran the simulation for $n = 500, 1000$, $T = 4, 8$, with 1000 Monte Carlo replications.

Table 1 reports the simulation results for the static logit model, based on CML estimates and Table 2 reports the results for the dynamic logit model, based on pseudo-CML estimates, with normally distributed covariates. The results for the design with χ^2_1 covariates are reported in Appendix in Tables 4 and 5. Each sub-panel of the tables reports the true value of the APE and

Table 1: Logit model (1), CML estimator

	$\beta_1 = 0, \beta_2 = 0$					$\beta_1 = 1, \beta_2 = -1$				
	True value	Mean bias	RMSE	Median bias	MAE	True value	Mean bias	RMSE	Median bias	MAE
$n = 500, T = 4$										
APE($\hat{\beta}_1$)	0.000	-0.000	0.011	-0.000	0.008	0.183	-0.004	0.012	-0.004	0.008
stderr[APE($\hat{\beta}_1$)]	0.011	0.000	0.001	0.000	0.000	0.011	-0.000	0.001	-0.000	0.000
APE($\hat{\beta}_2$)	0.000	-0.000	0.023	0.001	0.015	-0.186	0.004	0.023	0.004	0.016
stderr[APE($\hat{\beta}_2$)]	0.023	-0.000	0.001	-0.000	0.000	0.023	-0.000	0.001	-0.000	0.001
$n = 500, T = 8$										
APE($\hat{\beta}_1$)	0.000	-0.001	0.007	-0.000	0.005	0.183	-0.012	0.014	-0.012	0.012
stderr[APE($\hat{\beta}_1$)]	0.007	0.000	0.000	0.000	0.000	0.006	-0.000	0.000	-0.000	0.000
APE($\hat{\beta}_2$)	0.000	0.000	0.015	0.000	0.009	-0.187	0.013	0.019	0.013	0.014
stderr[APE($\hat{\beta}_2$)]	0.015	-0.000	0.000	-0.000	0.000	0.014	0.000	0.000	0.000	0.000
$n = 1000, T = 4$										
APE($\hat{\beta}_1$)	0.000	0.000	0.008	0.001	0.005	0.183	-0.004	0.009	-0.004	0.006
stderr[APE($\hat{\beta}_1$)]	0.008	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000
APE($\hat{\beta}_2$)	0.000	-0.000	0.017	-0.001	0.011	-0.186	0.004	0.016	0.004	0.011
stderr[APE($\hat{\beta}_2$)]	0.017	-0.001	0.001	-0.001	0.001	0.016	0.000	0.000	0.000	0.000
$n = 1000, T = 8$										
APE($\hat{\beta}_1$)	0.000	-0.000	0.005	0.000	0.003	0.183	-0.012	0.013	-0.012	0.012
stderr[APE($\hat{\beta}_1$)]	0.005	-0.000	0.000	-0.000	0.000	0.004	0.000	0.000	0.000	0.000
APE($\hat{\beta}_2$)	0.000	0.000	0.011	0.001	0.007	-0.187	0.013	0.016	0.013	0.013
stderr[APE($\hat{\beta}_2$)]	0.011	-0.000	0.000	-0.000	0.000	0.009	0.000	0.000	0.000	0.000

the standard deviation of its estimator in the simulation sample, the mean bias, the root mean square error (RMSE), the median bias, and the median absolute error (MAE) of the estimators of the APEs and of their standard errors.

The results in Table 1 suggest that the proposed two-step procedure provides a very close approximation of the true APE, with both the sample size and the time dimensions considered: when the true APEs depart from 0, the mean and median biases are rather small and the standard errors exhibit almost no bias. The results in Table 2 suggest that the estimator based on the pseudo-CML estimates of the dynamic logit regression parameters provides a good approximation of the true APEs as well. It is also worth noticing that while the RMSE and MAE of APE($\hat{\gamma}$) are slightly higher than the others with $T = 4$, they sharply decrease with $T = 8$.

4 Empirical application

We provide a real data application based on a dataset of unionised workers extracted form the US National Longitudinal Survey of Youth. The dataset is referred to 3648 women between 16 and 46 years old in 1968, interviewed between 1970 and 1988. The panel is unbalanced and the maximum number of occasions for the same subject is 12.

The response variable is the dummy "union", equal to 1 if the worker is unionised and 0

Table 2: Dynamic logit model (7), pseudo-CML estimator

	$\gamma = 0, \beta_1 = 1, \beta_2 = -2$					$\gamma = 1, \beta_1 = 1, \beta_2 = -2$				
	True value	Mean bias	RMSE	Median bias	MAE	True value	Mean bias	RMSE	Median bias	MAE
$n = 500, T = 4$										
APE($\hat{\gamma}$)	0.000	0.001	0.041	0.003	0.029	0.184	0.020	0.051	0.020	0.034
stderr[APE($\hat{\gamma}$)]	0.041	0.001	0.002	0.001	0.002	0.047	0.001	0.003	0.001	0.002
APE($\hat{\beta}_1$)	0.183	0.008	0.017	0.008	0.012	0.182	0.009	0.020	0.008	0.013
stderr[APE($\hat{\beta}_1$)]	0.041	0.001	0.002	0.001	0.002	0.047	0.001	0.003	0.001	0.002
APE($\hat{\beta}_2$)	-0.186	-0.005	0.030	-0.005	0.020	-0.185	-0.003	0.032	-0.004	0.021
stderr[APE($\hat{\beta}_2$)]	0.015	0.000	0.001	0.000	0.001	0.018	-0.001	0.001	-0.001	0.001
$n = 500, T = 8$										
APE($\hat{\gamma}$)	0.000	0.000	0.017	0.000	0.011	0.187	-0.010	0.021	-0.011	0.014
stderr[APE($\hat{\gamma}$)]	0.017	-0.001	0.001	-0.001	0.001	0.018	0.000	0.001	0.000	0.000
APE($\hat{\beta}_1$)	0.183	-0.011	0.013	-0.011	0.011	0.180	-0.015	0.016	-0.014	0.014
stderr[APE($\hat{\beta}_1$)]	0.017	-0.001	0.001	-0.001	0.001	0.018	0.000	0.001	0.000	0.000
APE($\hat{\beta}_2$)	-0.187	0.012	0.019	0.013	0.013	-0.187	0.019	0.024	0.020	0.020
stderr[APE($\hat{\beta}_2$)]	0.007	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000
$n = 1000, T = 4$										
APE($\hat{\gamma}$)	0.000	-0.000	0.030	-0.002	0.020	0.184	0.020	0.039	0.020	0.026
stderr[APE($\hat{\gamma}$)]	0.030	-0.000	0.001	-0.000	0.001	0.033	0.000	0.001	0.000	0.001
APE($\hat{\beta}_1$)	0.183	0.007	0.013	0.008	0.010	0.182	0.008	0.014	0.008	0.010
stderr[APE($\hat{\beta}_1$)]	0.030	-0.000	0.001	-0.000	0.001	0.033	0.000	0.001	0.000	0.001
APE($\hat{\beta}_2$)	-0.186	-0.005	0.023	-0.005	0.016	-0.185	-0.004	0.023	-0.005	0.016
stderr[APE($\hat{\beta}_2$)]	0.011	-0.001	0.001	-0.001	0.001	0.012	0.000	0.001	0.000	0.000
$n = 1000, T = 8$										
APE($\hat{\gamma}$)	0.000	0.000	0.012	0.000	0.008	0.187	-0.011	0.017	-0.011	0.012
stderr[APE($\hat{\gamma}$)]	0.012	-0.000	0.000	-0.000	0.000	0.013	-0.000	0.001	-0.000	0.000
APE($\hat{\beta}_1$)	0.183	-0.012	0.013	-0.012	0.012	0.180	-0.015	0.015	-0.015	0.015
stderr[APE($\hat{\beta}_1$)]	0.012	-0.000	0.000	-0.000	0.000	0.013	-0.000	0.001	-0.000	0.000
APE($\hat{\beta}_2$)	-0.187	0.013	0.016	0.013	0.013	-0.187	0.018	0.021	0.018	0.018
stderr[APE($\hat{\beta}_2$)]	0.005	-0.000	0.000	-0.000	0.000	0.005	-0.000	0.000	-0.000	0.000

Table 3: Estimation results for the static and dynamic logit model. Response variable: union. Regression coefficients and average partial effects

	Static logit		Dynamic logit	
	coeff.	APE	coeff.	APE
$union_{t-1}$			1.6104^{***}	0.3109^{***}
			[0.0757]	[0.0008]
age	0.0163^{***}	0.0028^{***}	0.0082^{*}	0.0013^{*}
	[0.0041]	[0.0011]	[0.0047]	[0.0008]
grade	0.0826^{**}	0.0142	-0.0252	-0.0040
	[0.0419]	[0.0101]	[0.0512]	[0.0086]
not smsa	-0.0274	-0.0047	0.0726	0.0115
	[0.1135]	[0.0279]	[0.1440]	[0.0241]
Log-likelihood	-4485.095		-3121.596	
# of subjects	1618		1393	

Standard errors in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.10. # of subjects refers to the actual number of workers contributing to the log-likelihoods in (5) and (11) for the static and dynamic logit, respectively.

otherwise. The regression covariates are two continuous variables "age" and "grade", which is the level of educational attainment, and the binary variable "not smsa", which is equal to 1 if the worker does not live in a metropolitan area and 0 otherwise.

Table 3 reports the estimation results for the regression coefficients and APEs for both the static and dynamic logit model, where the lagged dependent variable enters the set of covariates.

The estimation results for the static logit model show that age has a positive significant effect on the probability of being unionised, which amounts to around 0.3 percentage points for one more year. It is interesting to note that while the coefficient associated with "grade" is statistically significant at the 5% level, the results for the APE suggest that the level of educational attainment has not effect on the probability of being unionised. Finally, both the coefficient and APE associated with "not smsa" indicate that there is no effect of living in a metropolitan area on participating in unions. Looking at the estimation results for the dynamic logit model, a strong persistence in the dependent variable emerges, as indicated by the coefficient and the APE associated with the lagged response. The latter measures the effect of the true state dependence, which, in this case, entails that being unionised in $t-1$ increases the probability of being unionised at time t by 31 percentage points. With this specification, the only covariate still exerting a statistically significant effect is "age", whose effect however reduces to 0.1 percentage points.

5 Concluding remarks

Logit models that can be estimated by CML and PCML provide an attractive approach for the analysis of binary responses with fixed-effects, where incidental parameters are allowed to be correlated with the model covariates in a nonparametric way. As these estimators rely on eliminating the incidental parameters, the estimation of PEs and APEs, however, is not directly available.

We show that individual effects can be recovered by a simple additional step and then used to estimate PEs and APEs in a rather straightforward manner. We also provide a standard error formulation for the APEs that does not involve the estimators of the incidental parameters and therefore holds with fixed T, as also testified by our simulation results, thereby avoiding computationally intensive procedures such as bootstrapping. The real data application supports the key role of APEs in interpreting estimation results, especially because t-tests may yield contrasting results.

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

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 V. Nigro (2010): "A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator," *Econometrica*, 78, 719–733.

——— (2012): "Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data," *Journal of Econometrics*, 170, 102–116.

Bartolucci, F. and C. Pigini (2017): "cquad: An R and Stata Package for Conditional Maximum Likelihood Estimation of Dynamic Binary Panel Data Models," *Journal of Statistical Software*, 78.

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

——— (1985): "Heterogeneity, omitted variable bias, and duration dependence," in *Longitudinal analysis of labor market data*, ed. by J. J. Heckman and B. Singer, Cambridge University Press: Cambridge.

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

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

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

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

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

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

Appendix

Expressions for the derivatives in (12) are

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

and

$$\nabla_{\phi_k} m_i(\boldsymbol{\beta}, \phi_k) = -\frac{2}{T} \sum_t \left[f_{itk}(\boldsymbol{\beta}, \alpha_i(\boldsymbol{\beta}) - \phi_k \right].$$

The second derivatives in (14) 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

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

and $\boldsymbol{\nabla}_{\boldsymbol{\phi}\boldsymbol{\phi}} \boldsymbol{m}(\boldsymbol{\beta}, \boldsymbol{\phi})$ is a $K \times K$ diagonal matrix with element 2. Finally, for the computation of the block $\boldsymbol{\nabla}_{\boldsymbol{\phi}\boldsymbol{\beta}} m_i(\boldsymbol{\beta}, \boldsymbol{\phi})$ we rely on numerical differentiation.

Table 4: Logit model (1), CML estimator, χ_1^2 covariates

	$\beta_1 = 0, \beta_2 = 0$					$\beta_1 = 1, \beta_2 = -1$				
	True value	Mean bias	RMSE	Median bias	MAE	True value	Mean bias	RMSE	Median bias	MAE
$n = 500, T = 4$										
APE($\hat{\beta}_1$)	0.000	-0.000	0.012	-0.000	0.008	0.192	-0.004	0.017	-0.004	0.012
stderr[APE($\hat{\beta}_1$)]	0.012	-0.000	0.001	-0.000	0.000	0.017	0.000	0.001	0.000	0.001
APE($\hat{\beta}_2$)	0.000	-0.000	0.026	0.000	0.018	-0.185	0.001	0.026	0.002	0.017
stderr[APE($\hat{\beta}_2$)]	0.026	-0.000	0.001	-0.000	0.001	0.026	0.000	0.001	0.000	0.001
$n = 500, T = 8$										
APE($\hat{\beta}_1$)	0.000	-0.000	0.007	-0.000	0.005	0.182	-0.010	0.014	-0.011	0.011
stderr[APE($\hat{\beta}_1$)]	0.007	0.000	0.000	0.000	0.000	0.010	-0.000	0.001	-0.000	0.000
APE($\hat{\beta}_2$)	0.000	0.001	0.017	0.001	0.012	-0.173	0.008	0.017	0.008	0.012
stderr[APE($\hat{\beta}_2$)]	0.017	0.000	0.001	0.000	0.000	0.015	0.000	0.001	0.000	0.000
$n = 1000, T = 4$										
APE($\hat{\beta}_1$)	0.000	0.000	0.008	-0.000	0.005	0.192	-0.005	0.013	-0.005	0.009
stderr[APE($\hat{\beta}_1$)]	0.008	-0.000	0.000	-0.000	0.000	0.012	0.000	0.001	0.000	0.000
APE($\hat{\beta}_2$)	0.000	0.001	0.018	0.002	0.013	-0.185	0.002	0.020	0.001	0.014
stderr[APE($\hat{\beta}_2$)]	0.018	0.000	0.000	0.000	0.000	0.020	-0.001	0.001	-0.001	0.001
$n = 1000, T = 8$										
APE($\hat{\beta}_1$)	0.000	-0.000	0.005	-0.000	0.003	0.182	-0.011	0.013	-0.010	0.010
stderr[APE($\hat{\beta}_1$)]	0.005	0.000	0.000	0.000	0.000	0.007	-0.000	0.000	-0.000	0.000
APE($\hat{\beta}_2$)	0.000	-0.000	0.012	-0.000	0.008	-0.173	0.009	0.014	0.009	0.010
stderr[APE($\hat{\beta}_2$)]	0.012	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000

Table 5: Dynamic logit model (7), pseudo-CML estimator, χ_1^2 covariates

	$\gamma = 0, \beta_1 = 1, \beta_2 = -2$					$\gamma = 1, \beta_1 = 1, \beta_2 = -2$				
	True value	Mean bias	RMSE	Median bias	MAE	True value	Mean bias	RMSE	Median bias	MAE
$n = 500, T = 4$										
APE($\hat{\gamma}$)	0.000	0.001	0.044	0.000	0.028	0.185	0.019	0.051	0.020	0.036
stderr[APE($\hat{\gamma}$)]	0.044	-0.001	0.003	-0.001	0.002	0.048	0.000	0.003	0.000	0.002
APE($\hat{\beta}_1$)	0.180	0.011	0.019	0.010	0.012	0.181	0.011	0.021	0.010	0.013
stderr[APE($\hat{\beta}_1$)]	0.044	-0.001	0.003	-0.001	0.002	0.048	0.000	0.003	0.000	0.002
APE($\hat{\beta}_2$)	-0.187	-0.004	0.034	-0.005	0.024	-0.187	-0.003	0.037	-0.003	0.026
stderr[APE($\hat{\beta}_2$)]	0.015	0.000	0.001	0.000	0.001	0.018	-0.000	0.001	-0.001	0.001
$n = 500, T = 8$										
APE($\hat{\gamma}$)	0.000	-0.000	0.017	0.000	0.012	0.188	-0.011	0.021	-0.012	0.014
stderr[APE($\hat{\gamma}$)]	0.017	0.000	0.001	0.000	0.000	0.018	0.000	0.001	0.000	0.001
APE($\hat{\beta}_1$)	0.178	-0.009	0.011	-0.009	0.009	0.180	-0.013	0.015	-0.013	0.013
stderr[APE($\hat{\beta}_1$)]	0.017	0.000	0.001	0.000	0.000	0.018	0.000	0.001	0.000	0.001
APE($\hat{\beta}_2$)	-0.187	0.011	0.021	0.011	0.014	-0.187	0.019	0.025	0.018	0.019
stderr[APE($\hat{\beta}_2$)]	0.007	-0.000	0.000	-0.000	0.000	0.007	0.000	0.000	0.000	0.000
$n = 1000, T = 4$										
APE($\hat{\gamma}$)	0.000	-0.000	0.031	-0.001	0.021	0.185	0.019	0.038	0.017	0.025
stderr[APE($\hat{\gamma}$)]	0.031	0.000	0.001	0.000	0.001	0.034	0.000	0.001	0.000	0.001
APE($\hat{\beta}_1$)	0.180	0.011	0.015	0.011	0.011	0.181	0.009	0.015	0.009	0.011
stderr[APE($\hat{\beta}_1$)]	0.031	0.000	0.001	0.000	0.001	0.034	0.000	0.001	0.000	0.001
APE($\hat{\beta}_2$)	-0.187	-0.005	0.025	-0.005	0.017	-0.187	-0.003	0.025	-0.004	0.017
stderr[APE($\hat{\beta}_2$)]	0.010	0.000	0.001	0.000	0.000	0.012	0.001	0.001	0.001	0.001
$n = 1000, T = 8$										
APE($\hat{\gamma}$)	0.000	-0.000	0.012	-0.000	0.008	0.188	-0.011	0.017	-0.011	0.012
stderr[APE($\hat{\gamma}$)]	0.012	-0.000	0.000	-0.000	0.000	0.013	-0.000	0.000	-0.000	0.000
APE($\hat{\beta}_1$)	0.178	-0.009	0.010	-0.009	0.009	0.180	-0.013	0.014	-0.013	0.013
stderr[APE($\hat{\beta}_1$)]	0.012	-0.000	0.000	-0.000	0.000	0.013	-0.000	0.000	-0.000	0.000
APE($\hat{\beta}_2$)	-0.187	0.011	0.016	0.011	0.012	-0.187	0.019	0.023	0.020	0.020
stderr[APE($\hat{\beta}_2$)]	0.005	-0.000	0.000	-0.000	0.000	0.005	-0.000	0.000	-0.000	0.000