Modified profile likelihood for fixed-effects panel data models

F. Bartolucci[*], R. Bellio[†], A. Salvan[‡] and N. Sartori[‡]

September 6, 2013

Abstract

We show how modified profile likelihood methods, developed in the statistical literature, may be effectively applied to estimate the structural parameters of econometric models for panel data, with a remarkable reduction of bias with respect to ordinary likelihood methods. Initially, the implementation of these methods is illustrated for general models for panel data including individual-specific fixed effects and then, in more detail, for the truncated linear regression model and dynamic regression models for binary data formulated along with different specifications. Simulation studies show the good behavior of the inference based on the modified profile likelihood, even when compared to an ideal, although infeasible, procedure (in which the fixed effects are known) and also to alternative estimators existing in the econometric literature. The proposed estimation methods are implemented in an R package that we make available to the reader.

Keywords: autoregressive models; bias reduction; dynamic logit model; dynamic probit model; incidental parameter problem; truncated regression.

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

[†]Dipartimento di Scienze Economiche e Statistiche, Università di Udine, via Tomadini, 30, I-33100 Udine, Italy, ruggero.bellio@uniud.it

[‡]Dipartimento di Scienze Statistiche, Università di Padova, via C. Battisti, 241, I-35121 Padova, Italy, alessandra.salvan@unipd.it, nicola.sartori@unipd.it

1 Introduction

Panel data models usually include individual-specific fixed-effects parameters accounting for unobserved heterogeneity. Several solutions have been proposed to cope with the ensuing incidental parameters problem (Neyman and Scott, 1948). See Arellano and Hahn (2007) for a comprehensive review.

In estimating the structural parameters, which typically correspond to the parameters of interest, fixed effects may be removed when a marginal or conditional likelihood is available for these parameters (see e.g., Severini, 2000, Ch. 8). A well-known example is the conditional likelihood for the static logit model. However, the existence of these likelihoods depends on the assumptions of the model and is not guaranteed in general. For instance, a conditional likelihood is not available for the dynamic logit model, which includes the lagged response variable among the regressors, or for general probit models.

Several alternative solutions to the ones mentioned above have been developed in the literature; some of them are model-specific (e.g., Honoré and Kyriazidou, 2000, Bartolucci and Nigro, 2012) and some are of wider applicability. Here we focus on this second class of solutions. In their interesting review, Arellano and Hahn (2007) classify these general methods among: (*i*) those based on a bias correction of the maximum likelihood estimator, (*ii*) those based on a correction of the estimating equation, and (*iii*) those based on a correction of the target function, which typically is the profile (or concentrated) log likelihood function for the parameters of interest. Among the more recent contributions, we consider Fernández-Val (2009) and Fernández-Val and Vella (2011) for the first approach, Carro (2007) for the second, and Pace and Salvan (2006) and Bester and Hansen (2009) for the third. A common effect of the above corrections is that they lead to a reduction of the bias of the estimator of the parameters of interest from $O(T^{-1})$ to $O(T^{-2})$, where T is the number of time occasions.

Among the most effective solutions proposed in the statistical literature, there is the modified profile likelihood, which falls in class (*iii*) above and originates from the work of Barndorff-Nielsen (1980, 1983); see Severini (2000, Ch. 9) for an updated account. The properties of the modified profile likelihood and its approximations have been studied in detail by Sartori (2003) in a two-index asymptotic setting (in N, the sample size, and in T), which is a natural framework for panel data models. In particular, under fairly general conditions, a sufficient condition for the usual χ^2-asymptotic distribution of Wald, score, and likelihood ratio statistics is that $N = o(T^3)$, whereas it is that $N = o(T)$ for the profile likelihood.

Aim of the present paper is to show how a convenient version of the modified profile likelihood (Severini, 1998) may be effectively applied to estimate fixed-effects versions of some important econometric models for panel data. In particular, we illustrate its use in truncated linear regression and in dynamic logit and probit models, and we provide an assessment of its properties through a series of simulations. An `R` code for the methods used in this paper is included in the package `panelMPL`, which is available at `http://www.dies.uniud.it/index.php/software-bellio.html`.

The paper is organized as follows. In the next section we describe the modified profile likelihood, which is then applied to some relevant cases in Section 3. Some final remarks are given in Section 4.

2 Modified profile likelihood

2.1 Background and notation

With reference to a sample of N units observed at T occasions, let y_{it} denote the response variable and let x_{it} denote the column vector of covariates observed for sample unit i at occasion t, with $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The covariates are considered strictly exogenous, with a possible exception being that of the lag of the endogenous variable included among the covariates. Let also $f(y_{it}|x_{it}; \theta, \alpha_i)$ denote the probability mass or density function of y_{it} given x_{it}. It is assumed that, given x_{it}, y_{it} is conditionally independent of $y_{i1}, \ldots, y_{i,t-1}$, or it is independent of $y_{i1}, \ldots, y_{i,t-h}$ for a suitable h for dynamic models in which $y_{i,t-h+1}, \ldots, y_{i,t-1}$ are included in x_{it}. The vector of structural parameters is denoted by θ, whereas $\alpha_1, \ldots, \alpha_N$ are, without loss of generality, scalar incidental parameters.

Likelihood inference about θ is based on the profile, or concentrated, log likelihood

$$\ell_P(\theta) = \sum_i \ell_i(\theta, \hat{\alpha}_i(\theta)),$$

where $\ell_i(\theta, \alpha_i) = \sum_t \log f(y_{it}|x_{it}; \theta, \alpha_i)$ is the log likelihood contribution for sample unit i, $i = 1, \ldots, N$, and $\hat{\alpha}_i(\theta)$ is the constrained maximum likelihood estimate of α_i for fixed θ, that is,

$$\hat{\alpha}_i(\theta) = \arg\max_{\alpha_i} \ell_i(\theta, \alpha_i),$$

which is typically the solution of $\ell_{\alpha_i}(\theta, \alpha_i) = \partial\ell(\theta, \alpha_i)/\partial\alpha_i = \partial\ell_i(\theta, \alpha_i)/\partial\alpha_i = 0$. Throughout the paper we use the convention that when the range of the sum is not explicitly indicated it is equal to $1, \ldots, N$ (for the sums in i) or it is equal to $1, \ldots, T$ (for the sums in t). Note that maximizing the profile log likelihood $\ell_P(\theta)$ leads to the same solution as maximizing the overall log likelihood, that is,

$$\ell(\theta, \alpha) = \sum_i \ell_i(\theta, \alpha_i),$$

with respect to θ and α, where α is the vector of all individual-effects, with elements $\alpha_1, \ldots, \alpha_N$.

Typically, maximization of $\ell_P(\theta)$ is performed numerically, obtaining in this way the maximum likelihood estimator $\hat{\theta}$; see, for instance, Greene (2004). Moreover, standard output of optimization routines provides the Hessian of $\ell_P(\theta)$ at $\hat{\theta}$, which produces asymptotically correct estimates of the standard errors.

2.2 Severini (1998)'s modified profile likelihood

The maximum likelihood estimator $\hat{\theta}$ is in general asymptotically biased, with bias of order $O(T^{-1})$. A substantial reduction of this bias may be obtained by the estimator which maximizes the modified profile log likelihood, which has the general form

$$\ell_M(\theta) = \ell_P(\theta) + M(\theta), \tag{1}$$

where, because of the independence of the sample units, the adjustment function $M(\theta)$ has the additive form

$$M(\theta) = \sum_i M_i(\theta). \tag{2}$$

Each term $M_i(\theta)$ is of order $O_p(1)$ under repeated sampling, so that the total adjustment to the profile log likelihood is of order $O_p(N)$ when both T and N tend to infinity. We will denote by $\hat{\theta}_M$ the estimator based on the maximization of $\ell_M(\theta)$.

The first version of (1) that has been developed is the modified profile log likelihood of Barndorff-Nielsen (1980, 1983), introduced with the aim of obtaining an accurate approximation of a marginal or a conditional log likelihood, when either exists. The computation of this version of the modified profile likelihood requires the specification of an exact, or approximate, ancillary statistic; see Severini (2000, Ch. 6). This is usually straightforward in full exponential or composite group families. See Brazzale et al. (2007) for an overview.

The drawback of the original formulation of the modified profile likelihood is that it may be difficult to compute in most models which do not belong to full exponential or composite group families. However, several approximations are available which maintain the same asymptotic properties; for a review, see Severini (2000, Sec. 9.3). These approximations are to a large extent equivalent in terms of accuracy.

When θ and α are orthogonal, that is, when the corresponding block of the expected information matrix has all elements equal to 0, the generic term of the modification (2) can be taken equal to

$$M_i(\theta) = -\frac{1}{2} \log \mid -\ell_{\alpha_i \alpha_i}(\theta, \hat{\alpha}_i(\theta)) \mid .$$

This modification gives the approximate conditional likelihood of Cox and Reid (1987); see Lancaster (2002) for applications to panel data models. Although leading to a strong simplification, an orthogonal parameterization may be difficult to find or even it could not exist when θ is multidimensional. Moreover, the Cox and Reid approximate conditional likelihood is not invariant under interest-preserving reparameterizations, that is, under one-to-one functions from (θ, α) to $(\phi(\theta), \psi(\theta, \alpha))$.

Different approximations of the modified profile log likelihood are available which overcome these difficulties and avoid the specification of an ancillary statistic (see Severini, 2000, Sec. 6.7, for a review). Here, in particular, we focus on the version proposed by Severini (1998), which relies on modification (2) with

$$M_i(\theta) = \frac{1}{2}\log\mid -\ell_{\alpha_i\alpha_i}(\theta,\hat{\alpha}_i(\theta))\mid - \log\mid I_{\alpha_i\alpha_i}(\hat{\theta},\hat{\alpha}_i;\theta,\hat{\alpha}_i(\theta))\mid, \tag{3}$$

where

$$I_{\alpha_i\alpha_i}(\theta_0,\alpha_{i0};\theta_1,\alpha_{i1}) = E_{\theta_0,\alpha_{i0}}\left(\ell_{\alpha_i}(\theta_0,\alpha_{i0})\ell_{\alpha_i}(\theta_1,\alpha_{i1})\right). \tag{4}$$

Example 1 *First order autoregressive model.* Consider the model

$$y_{it} - \mu_i = (y_{i,t-1} - \mu_i)\rho + \varepsilon_{it}\sigma, \qquad i = 1,\ldots,N, \quad t = 2,\ldots,T, \tag{5}$$

with the error terms ε_{it} independent and normally distributed with mean zero and variance σ^2, and with $\rho \in (-1,1)$ and $\mu_i \in \mathbb{R}$, $i = 1,\ldots,N$. Note that, when $\rho = 0$, this model simplifies to the classic example of Neyman and Scott (1948). As for the initial condition, we assume that

$$y_{i1} \sim N\left(\mu_i, \frac{\sigma^2}{1-\rho^2}\right).$$

Equation (5) can also be written in the form

$$y_{it} = \delta_i + y_{i,t-1}\rho + \varepsilon_{it}\sigma, \tag{6}$$

where $\delta_i = \mu_i(1-\rho)$ is an individual-specific intercept. Specifications (5) and (6) are equivalent for the inference problems considered here. In particular, we consider $\theta = (\rho,\sigma^2)$ as the structural parameter, whereas the vector of individual-specific effects $\alpha = \mu$, with $\mu = (\mu_1,\ldots,\mu_N)$, will be treated as a nuisance parameter.

Profile likelihood and modifications are better expressed using matrix notation. Therefore, we let $y_i = (y_{i1},\ldots,y_{iT})^\top$, with $y_i \sim N_T(\alpha_i 1_T, \Omega)$, where 1_T is a vector of ones of length T and the matrix $\Omega = \Omega(\theta)$ has generic element $\omega_{tu} = \sigma^2\rho^{|t-u|}/(1-\rho^2)$, with

$t, u = 1, \ldots, T$. The inverse of Ω is

$$\Omega^{-1} = \sigma^{-2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

Then, we have

$$\ell_i(\theta, \alpha_i) = -\frac{1}{2} \log |\Omega| - \frac{1}{2} (y_i - \alpha_i 1_T)^\top \Omega^{-1} (y_i - \alpha_i 1_T)\,,$$

where $|\Omega| = (\sigma^2)^T/(1-\rho^2)$, and

$$\ell_{\alpha_i}(\theta, \alpha_i) = (y_i - \alpha_i 1_T)^\top \Omega^{-1} 1_T\,,$$

so that

$$\hat{\alpha}_i(\theta) = \frac{y_i^\top \Omega^{-1} 1_T}{1_T^\top \Omega^{-1} 1_T}\,.$$

The contribution of the ith unit to $\ell_P(\theta)$ is given by

$$\ell_i(\theta, \hat{\alpha}_i(\theta)) = -\frac{1}{2} \log |\Omega| - \frac{1}{2} y_i^\top \Psi y_i\,,$$

with

$$\Psi = \Omega^{-1} \left[I_T - 1_T (1_T^\top \Omega^{-1} 1_T)^{-1} 1_T^\top \Omega^{-1} \right],$$

and I_T being the identity matrix of order T.

It is straightforward to verify that

$$-\ell_{\alpha_i\alpha_i}(\theta, \hat{\alpha}_i(\theta)) = I_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta)) = 1_T^\top \Omega^{-1} 1_T$$

giving

$$\ell_M(\theta) = \ell_P(\theta) - \frac{N}{2} \log |1_T^\top \Omega^{-1} 1_T| \,.$$

Note that α and θ orthogonal and in this case $\ell_M(\theta)$ coincides with the approximate conditional log likelihood of Cox and Reid (1987).

A marginal likelihood for θ is available for this model, see Cruddas, Reid and Cox (1989) and the references therein. Its ith contribution is proportional to the marginal density of $d_i = y_i - 1_T \bar{y}_i$, with $\bar{y}_i = T^{-1} 1_T^\top y_i$, and its expression is most easily obtained as

$$L_{marg}(\theta; d_i) = \int_{-\infty}^{+\infty} L_i(\theta, \alpha_i) d\alpha_i \,;$$

see, for instance, Barndorff-Nielsen and Cox (1994), Sec. 2.8. It is easily checked that, for this model, $\sum_i \log L_{marg}(\theta; d_i)$ is equivalent to $\ell_M(\theta)$, so that the latter provides a consistent estimator as N diverges, even with fixed T. On the contrary, the profile likelihood has the usual problems, especially related to the bias and the ensuing poor coverage of confidence intervals, as proved by simulation results that are not reported here; these results also confirm the extremely accurate behavior of $\ell_M(\theta)$.

Covariates may be added to the model by considering $y_i \sim N_T(\mu_i 1_T + X_i \beta, \Omega)$, with X_i being a covariate matrix with T rows, still giving that $\ell_M(\theta)$ is equivalent to a marginal likelihood, where now θ also includes β. □

2.3 Computational aspects

Typically, the estimate $\hat{\theta}_M$ which maximizes $\ell_M(\theta)$ is obtained through numerical optimization. Evaluation of standard errors is obtained by using the second derivative of $\ell_M(\theta)$ at the maximum, which is a standard output of most numerical algorithms. This is what we implemented in the `R` package `panelMPL`. The estimate $\hat{\theta}_M$ and the corresponding estimated standard error can be used to compute Wald confidence intervals, based on asymptotic normality of the estimator. In general, intervals based on the likelihood ratio statistic constructed using $\ell_M(\theta)$ would be preferable for various reasons, including equivariance under reparameterizations. However, given that the information is of order $O(NT)$, for large N, both profile and modified profile likelihood are well approximated by a quadratic function in a neighborhood of the maximum and, therefore, Wald confidence intervals are almost indistinguishable from those based on the likelihood ratio statistic.

Regarding the computation of Severini (1998)'s version of $\ell_M(\theta)$, analytical calculation of (4) is fairly simple in a number of important models as those in Example 1 and Section 3.1; see also Bellio and Sartori (2006). Exact computation by enumeration is feasible in discrete models for moderate T, while Monte Carlo calculation offers a general solution (Pierce and Bellio, 2006); both alternatives are used in the following Section 3.2.

If observations within units are independent, as in static models, an asymptotically equivalent version of (3) is obtained by replacing $I_{\alpha_i\alpha_i}(\hat\theta, \hat\alpha_i; \theta, \hat\alpha_i(\theta))$ by its empirical analogue

$$\hat I_{\alpha_i\alpha_i}(\hat\theta, \hat\alpha_i; \theta, \hat\alpha_i(\theta)) = \sum_t \frac{\partial \log f(y_{it}|x_{it}; \hat\theta, \hat\alpha_i)}{\partial \alpha_i} \frac{\partial \log f(y_{it}|x_{it}; \theta, \hat\alpha_i(\theta))}{\partial \alpha_i};$$

see Severini (1999; 2000, Sec. 9.5.5). A similar empirical approximation is considered by Di Ciccio and Stern (1993) and DiCiccio *et al.* (1996); see also Arellano and Hahn (2007) that, apart a constant term, found the following expression

$$M_i(\theta) = \frac{1}{2} \log | -\ell_{\alpha_i\alpha_i}(\theta, \hat\alpha_i(\theta)) | - \frac{1}{2} \log \sum_t \left[\frac{\partial \log f(y_{it}|x_{it}; \theta, \hat\alpha_i(\theta))}{\partial \alpha_i} \right]^2 .$$

However, adjustments involving model based quantities are typically superior to their empirical counterparts, in particular when the unit sample size is small (see, for instance, Severini, 1999, and Bester and Hansen, 2009).

3 Examples

We illustrate the application of the methods outlined in Section 2 to some models of interest. In particular, we consider the truncated linear regression model and dynamic regression models for binary data. We also show the results of a set of simulations performed, under different scenarios, along the same lines as in Honoré and Kyriazidou (2000) and Carro (2007) among others. Based on these simulation results, the performance of the estimators of the parameters of interest are measured in terms of bias (B), median bias (MB), root mean squared error (RMSE), and median absolute error (MAE). For these estimators, we also report the standard deviation (SD) and the ratio SE/SD, where SE stands for the average over simulations of likelihood based estimates of the standard error, and the empirical coverage of 0.95 confidence intervals constructed assuming the asymptotic normality of the corresponding estimator. Moreover, in order to measure the actual improvement of the modified profile likelihood estimator (MPL estimator for short) over the estimator based on the standard profile likelihood (ML estimator), we compute for each simulation scenario the index

$$\Delta(\mathrm{B}) = \frac{|\mathrm{B(ML)}| - |\mathrm{B(MPL)}|}{|\mathrm{B(ML)}| - |\mathrm{B(IL)}|}, \tag{7}$$

where B(IL) stands for the bias of the infeasible likelihood estimator (IL estimator), which uses the fixed effects as one of the explanatory variables and treats the corresponding regression coefficient as unknown; see Honoré and Kyriazidou (2000). Indices analogous to (7) are also computed for MB, RMSE, and MAE and are denoted by Δ(MB), Δ(RMSE), and Δ(MAE), respectively. An additional comparison with other proposals in the literature is considered for the dynamic logit model in Section 3.2.

3.1 Truncated linear regression model

Let y_{it} be distributed as y^*_{it} conditionally on $y^*_{it} > 0$, with

$$y^*_{it} = \eta_{it} + \varepsilon_{it}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{8}$$

where $\eta_{it} = \eta_{it}(\alpha_i, \beta)$ is a linear predictor of the form

$$\eta_{it} = \alpha_i + x_{it}^\top \beta, \qquad i = 1, \dots, N, \quad t = 1, \dots, T,$$

with parameters $\alpha_i \in \mathbb{R}$, $i = 1, \dots, N$, $\beta \in \mathbb{R}^p$. The errors ε_{it} are assumed to be independent and normally distributed with mean zero and variance σ^2. Then, y_{it} has density

$$f(y_{it}|x_{it}; \beta, \sigma, \alpha_i) = \frac{1}{\sigma}\, \phi\left(\frac{y_{it} - \eta_{it}}{\sigma}\right) \Big/ \, \Phi\left(\frac{\eta_{it}}{\sigma}\right),$$

where $\phi(\cdot)$ is the density function of the standard normal distribution and $\Phi(\cdot)$ is the corresponding cumulative distribution function. We consider $\theta = (\beta^\top, \sigma)$ as the parameter of interest, whereas $\alpha = (\alpha_1, \ldots, \alpha_N)$ is treated as a vector of nuisance parameters.

The full log likelihood for model (8) is

$$\ell(\theta, \alpha) = -NT \log \sigma - \frac{1}{2\sigma^2} \sum_i \sum_t (y_{it} - \eta_{it})^2 - \sum_i \sum_t \log \Phi\left(\frac{\eta_{it}}{\sigma}\right);$$

the profile log likelihood for θ has the same expression with every α_i substituted by $\hat{\alpha}_i(\theta)$, which is the solution with respect to α_i of the equation

$$\frac{1}{\sigma} \sum_t (y_{it} - \eta_{it}) = \sum_t \phi\left(\frac{\eta_{it}}{\sigma}\right) \Big/ \Phi\left(\frac{\eta_{it}}{\sigma}\right).$$

Simulation results in Greene (2004, Sec. 3.2), with $N = 1000$ and T ranging from 2 to 20, show that $\hat{\theta}$ has a non-negligible negative bias. In particular, negative biases are accompanied by extremely poor coverage of the confidence intervals. These results are confirmed by our simulation study described below.

We compute the modified profile log likelihood given by (1) and (2) using expression (3) for $M_i(\theta)$. The quantities needed for the adjustment term $M_i(\theta)$ are

$$\begin{aligned}
\ell_{\alpha_i\alpha_i}(\theta, \hat{\alpha}_i(\theta)) &= -\frac{1}{\sigma^2}\left[T - \sum_t \left(\frac{\phi(\tilde{\eta}_{it}/\sigma)}{\Phi(\tilde{\eta}_{it}/\sigma)}\right)^2 - \frac{1}{\sigma}\sum_t \frac{\phi(\tilde{\eta}_{it}/\sigma)}{\Phi(\tilde{\eta}_{it}/\sigma)}\tilde{\eta}_{it}\right], \\
I_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta)) &= \frac{1}{\sigma^2}\left[T - \sum_t \left(\frac{\phi(\hat{\eta}_{it}/\hat{\sigma})}{\Phi(\hat{\eta}_{it}/\hat{\sigma})}\right)^2 - \frac{1}{\hat{\sigma}}\sum_t \frac{\phi(\hat{\eta}_{it}/\hat{\sigma})}{\Phi(\hat{\eta}_{it}/\hat{\sigma})}\hat{\eta}_{it}\right],
\end{aligned}$$

where $\tilde{\eta}_{it} = \eta_{it}(\hat{\alpha}_i(\theta), \beta)$ and $\hat{\eta}_{it} = \eta_{it}(\hat{\alpha}_i, \hat{\beta})$. Thus, discarding additive terms that do not depend on θ, we have

$$\ell_M(\theta) = \ell_P(\theta) + N \log \sigma + \frac{1}{2}\sum_i \log\left[T - \sum_t \left(\frac{\phi(\tilde{\eta}_{it}/\sigma)}{\Phi(\tilde{\eta}_{it}/\sigma)}\right)^2 - \frac{1}{\sigma}\sum_t \frac{\phi(\tilde{\eta}_{it}/\sigma)}{\Phi(\tilde{\eta}_{it}/\sigma)}\tilde{\eta}_{it}\right].$$

In order to assess the accuracy of the inference based on $\ell_M(\theta)$ in short panels with many sample units, we implement a simulation experiment based on 5000 random samples with $N = 250, 500, 1000$ and $T = 4, 8$. Under this model, $\beta = \sigma = 1$, whereas, for each pair

(T, N), covariates x_{it} are generated from a standard normal distribution and the values of the incidental parameters α_i are chosen as follows:

$$\alpha_i = \sqrt{T}\bar{x}_i + u_i, \quad i = 1, \ldots, N,$$

where $\bar{x}_i = T^{-1}\sum_t x_{it}$ and $u_i \sim N(0, 1)$.

Tables 1 and 2 report the results for β and $\omega = \log\sigma$, respectively. The latter parameterization was chosen for numerical convenience.

[Table 1 about here.]

[Table 2 about here.]

Bias and median bias are approximately equal in all cases and MPL gives a substantial improvement over ML. As expected, the bias decreases as T increases, whereas it does not depend on N. On the other hand, the root mean square error depends both on T and on N. Again, this was expected, since the variance of the estimator is of order $O(1/(NT))$. Notice that all the Δ indices are larger than 0.73, and often larger than 0.90. For instance, with $N = 1000$ and $T = 8$ we get $\Delta(\text{RMSE}) = 0.898$, implying that MPL produces about 90% of the RMSE reduction from ML to the infeasible estimator. Coverage properties of MPL confidence intervals are excellent. The improvement over ML is remarkable and this is largely due to bias reduction. There is also a curvature correction, being SE/SD for MPL much closer to one than for ML. Unreported simulation results confirm that confidence intervals based on the likelihood ratio statistics are indistinguishable from Wald confidence intervals, as commented in Section 2.3.

3.2 Dynamic regression models for binary data

Let y_{it} be observations from Bernoulli variables with success probabilities $G(\eta_{it})$, where $G(\cdot)$ is a given continuous cumulative distribution function on $\mathbb{R}$ with density $g(\cdot)$ and η_{it} is a linear predictor including a dynamic effect, that is,

$$\eta_{it} = \alpha_i + x_{it}^\top\beta + y_{it-1}\rho, \qquad i = 1, \ldots, N, \quad t = 2, \ldots, T, \tag{9}$$

with parameters $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, N$, $\beta \in \mathbb{R}^p$, and $\rho \in \mathbb{R}$. Typical choices of $G(\cdot)$ are the logistic distribution and the normal distribution, leading to the logistic regression model and the probit regression model, respectively. We assume that the initial observation y_{i1} is a fixed non-stochastic constant for unit i, as in Carro (2007) and Fernández-Val (2009), among others.

Honoré and Kyriazidou (2000) proposed an estimator that for $T = 4$ is based on a particular weighted likelihood, whereas for $T > 4$ it is based on a pairwise weighted likelihood. In either case, their solution provides a consistent estimator but it has some limitations, such as the impossibility of estimating the effects of time dummies or other categorical covariates. See Bartolucci and Nigro (2010, 2012) for more specific comments and alternative approaches based on the conditional likelihood. Hahn and Kuersteiner (2011) and Fernández-Val (2009) provided formulas for removing the $O(T^{-1})$ leading term of the bias of the maximum likelihood estimator. Carro (2007) removed the leading term of the bias of the profile score for $\theta = (\beta^\top, \rho)$. The dynamic logit model is also one of the examples considered by Bester and Hansen (2009) for their approach based on penalization.

Under assumption (9), the profile log likelihood for $\theta = (\beta^\top, \rho)$ is

$$\ell_P(\theta) = \sum_i \sum_{t>1} \left[y_{it} \log G(\tilde{\eta}_{it}) + (1 - y_{it}) \log\left(1 - G(\tilde{\eta}_{it})\right) \right],$$

where $\tilde{\eta}_{it} = \hat{\alpha}_i(\theta) + x_{it}^\top \beta + y_{it-1}\rho$ is the constrained maximum likelihood estimate of η_{it} for the ith unit.

The computation of the adjustment terms $M_i(\theta)$ given by (3) requires

$$-\ell_{\alpha_i\alpha_i}(\theta, \hat{\alpha}_i(\theta)) = \sum_{t>1} \left[\frac{g(\tilde{\eta}_{it})^2}{G(\tilde{\eta}_{it})\,(1 - G(\tilde{\eta}_{it}))} - C(\tilde{\eta}_{it}) \right],$$

with

$$C(\tilde{\eta}_{it}) = (y_{it} - G(\tilde{\eta}_{it})) \left[\frac{g'(\tilde{\eta}_{it})}{G(\tilde{\eta}_{it})\,(1 - G(\tilde{\eta}_{it}))} - \frac{g(\tilde{\eta}_{it})^2\,(1 - 2G(\tilde{\eta}_{it}))}{G(\tilde{\eta}_{it})^2\,(1 - G(\tilde{\eta}_{it}))^2} \right]$$

and $I_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta))$, which does not have a closed-form expression and cannot be computed by the recursive computation employed by Carro (2007). A feasible approach is based on the fact that the sample space $\mathcal{Y}$ of all the possible vectors $(y_{i2}, \ldots, y_{iT})$ has cardinality 2^{T-1}. Therefore, for moderate values of T it is possible to explore all $\mathcal{Y}$ for computing $I_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta))$. In other words, letting $s = (s_2, \ldots, s_T) \in \mathcal{Y}$, we compute

$$\begin{aligned} p(s; \hat{\theta}, \hat{\alpha}_i) &= p(s_2, \ldots, s_T; \hat{\theta}, \hat{\alpha}_i) = \prod_{t>1} G(\hat{\eta}_{it})^{s_t}(1 - G(\hat{\eta}_{it}))^{1-s_t}, \\ I_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta)) &= \sum_{s \in \mathcal{Y}} \left[\sum_{t>1} \frac{g(\hat{\eta}_{it})(s_t - G(\hat{\eta}_{it}))}{G(\hat{\eta}_{it})(1 - G(\hat{\eta}_{it}))} \right] \left[\sum_{t>1} \frac{g(\tilde{\eta}_{it})(s_t - G(\tilde{\eta}_{it}))}{G(\tilde{\eta}_{it})(1 - G(\tilde{\eta}_{it}))} \right] p(s; \hat{\theta}, \hat{\alpha}_i)\,. \end{aligned}$$

For $t > 2$, the estimated linear predictors are $\hat{\eta}_{it} = \hat{\alpha}_i + x_{it}^\top \hat{\beta} + \hat{\rho} s_{t-1}$ and $\tilde{\eta}_{it} = \hat{\alpha}_i(\theta) + x_{it}^\top \beta + \rho s_{t-1}$, while for $t = 2$ they are $\hat{\eta}_{i2} = \hat{\alpha}_i + x_{i2}^\top \hat{\beta} + \hat{\rho} y_{i1}$ and $\tilde{\eta}_{i2} = \hat{\alpha}_i(\theta) + x_{i2}^\top \beta + \rho y_{i1}$. With larger values of T, say $T > 10$, full exploration of $\mathcal{Y}$ may become computationally costly. An alternative strategy relies on a Monte Carlo approximation, that is, $I_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta)) \doteq I^*_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta))$, with

$$I^*_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta)) = \frac{1}{R} \sum_{r=1}^{R} \left[\sum_{t>1} \frac{g(\hat{\eta}^*_{it})(y^*_{r,it} - G(\hat{\eta}^*_{it}))}{G(\hat{\eta}^*_{it})(1 - G(\hat{\eta}^*_{it}))} \right] \left[\sum_{t>1} \frac{g(\tilde{\eta}^*_{it})(y^*_{r,it} - G(\tilde{\eta}^*_{it}))}{G(\tilde{\eta}^*_{it})(1 - G(\tilde{\eta}^*_{it}))} \right],$$

where $y^*_{r,i} = (y_{r,i2}, \ldots, y_{r,iT})$, $r = 1, \ldots, R$, is generated from the model with $\theta = \hat{\theta}$ and $\alpha_i = \hat{\alpha}_i$. The linear predictors $\hat{\eta}^*_{it}$ and $\tilde{\eta}^*_{it}$ involve the simulated data as lagged variables, but are otherwise computed using the parameter estimates $(\hat{\theta}^\top, \hat{\alpha}_i)$ and $\hat{\alpha}_i(\theta)$, respectively. The computation of $I^*_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta))$ only requires the score function $\ell_{\alpha_i}(\theta, \alpha_i)$ and simulation from the model, without any additional fitting; therefore, it is easily applicable in wide generality. This solution, used for the results in Tables 7 and 8, has also the advantage of having a moderate computational cost, which is only slightly affected by the value of T. In our experience, values of R even lower than 1000 may be sufficient for reasonable accuracy in most cases, though in specific applications this fact may require some trial and error with more than one value of R. We note that computation of $\ell_M(\theta)$ using $I^*_{\alpha_i\alpha_i}(\hat{\theta}, \hat{\alpha}_i; \theta, \hat{\alpha}_i(\theta))$ is only partially akin to simulated maximum likelihood, as Monte Carlo computation is used for computing part of the adjustment term, rather than the entire likelihood function. Nonetheless, it is advisable to generate observations employing the same set of random draws during the maximization of $\ell_M(\theta)$, as customary in simulated maximum likelihood (i.e., Hajivassiliou, 2000).

Tables 3–6 summarize results for the logit and probit models in simulations where $x_{it} \sim N(0, \pi^2/3)$ and $\alpha_i = T^{-1}\sum_{i=1}^{T} x_{it}$, with the inclusion of the lagged response among the covariates. The regression parameter in the probit model has been set equal to that of the logit parameter divided by 1.6, in order to make the two regression functions approximately equivalent. The initial conditions y_{i1} were generated using the linear predictor (9) with $\rho = 0$. As noted by Carro (2007), and analogously to the static case, units with no variation in the response do not contribute to the profile or the modified profile likelihood, hence the actual sample size is reduced.

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

The relative behavior of the profile and modified profile likelihood inference is quite similar to that observed in the simulation study for the truncated regression model and most of the comments to Tables 1 and 2 apply here as well, with Δ indices close to 0.90 or even higher. However, the finite sample properties of the maximum likelihood estimator are so poor that the adjustment provided by $\ell_M(\theta)$ is at times only partially successful, that is, the $O(T^{-2})$ bias of $\hat{\theta}_M$ can be at times not negligible and the curvature correction is less good than in the static case ($\rho = 0$), simulation results for which are not reported here. This fact can be observed in some of the entries of Tables 3–6, with notable undercoverage particularly for $T = 4$. At any rate, the results based on $\ell_M(\theta)$ are in line with the best proposals in the literature. For the logit case, which is the most analyzed by other authors, the setting of this paper is very similar to that in Honoré and Kyriazidou (2000), Carro (2007), Bester and Hansen (2009), Fernández-Val (2009), and Hahn and Kuersteiner (2011). In particular, all these studies report the results for $T = 8$ and $N = 250$. Among the more general large T-consistent estimators, the solution proposed here is comparable with the best performing ones, that are Carro (2007) (MML) and Fernández-Val (2009) (BC), which are also included in the simulation studies summarized in Tables 3 and 4. With $T = 4$, MPL outperforms the other methods, with BC showing a very poor behavior. On the other hand, when $T = 8$, BC is quite good, even slightly better than MPL. Both are preferable to MML.

Finally, Tables 7 and 8 show a comparison for $T = 11$ and $T = 16$ of MPL with BC and MML. Here, computation of (4) needed for MPL has been performed using Monte Carlo with $R = 500$ replications. We note that all methods are quite accurate in all cases, giving very small biases and empirical coverages of confidence intervals close to the nominal level. As a check of the accuracy of the Monte Carlo version of MPL we also ran simulations in the case $T = 8$. Results are almost indistinguishable from those of the exact version reported in Tables 3 and 4, and therefore are not reported here.

[Table 7 about here.]

[Table 8 about here.]

4 Conclusions

We study the application of modified profile likelihood methods to econometric models for panel data. These methods have been mainly developed in the statistical literature, starting from the fundamental work of Barndorff-Nielsen (1980, 1983), and are specially tailored to deal with models with incidental parameters (Sartori, 2003); see also Severini (2000, Ch. 9). Thus, we establish a bridge between the statistical and the econometric literature regarding inference for incidental parameters models. Modified profile likelihood is not only general and interpretable, but also quite effective for inference about structural parameters. Our results also indicate that it is highly competitive with existing estimation methods developed in the econometric literature.

The implementation of the modified profile likelihood in general models can be quite straightforward, in particular if a Monte Carlo version is considered to compute certain quantities required for the implementation. An R package `panelMPL` for the models used in the paper is available to the reader.

In this paper, we focus on estimation of structural parameters, while we have not explicitly considered marginal or partial effects. These are typically function of all parameters in the model. One way to estimate these effects is to use $\hat{\theta}_M$ and $\hat{\alpha}_1(\hat{\theta}_M), \dots, \hat{\alpha}_N(\hat{\theta}_M)$, where, as before, $\hat{\alpha}_i(\theta)$ is the constrained maximum likelihood estimate of α_i for fixed θ. This is the same approach considered in Carro (2007), which proved to be quite satisfactory.

Finally, we note that a possible extension to estimating equations could be obtained using the approach of Severini (2002), as done in the context of generalized estimating equations by Wang and Hanfelt (2008).

Acknowledgement

Francesco Bartolucci acknowledges the financial support from the "Einaudi Institute for Economics and Finance" EIEF (Rome - IT) and the Italian Government (FIRB - "Futuro in Ricerca"). Nicola Sartori was supported by the Cariparo Foundation Excellence grant 2011/2012. Ruggero Bellio, Alessandra Salvan and Nicola Sartori were supported by MIUR PRIN 2008. We thank the reviewers for their constructive comments.

References

Arellano, M., Hahn, J. (2007). Understanding bias in nonlinear panel models: some recent developments. In: R. Blundell, W. Newey, T. Persson (Eds) *Advances in Economics and Econometrics*, ninth world congress, Vol.3, Ch.12. Cambridge University Press, Cambridge.

Barndorff-Nielsen, O.E. (1980). Conditionality resolutions. *Biometrika*, **67**, 293–310.

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

Barndorff-Nielsen, O.E., Cox, D.R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

Bartolucci, F., Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a root-n consistent conditional estimator. *Econometrica*, **78**, 719–733.

Bartolucci, F., Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data, *Journal of Econometrics*, **170**, 102–116.

Bellio, R., Sartori, N. (2006). Practical use of modified maximum likelihoods for stratified data. *Biometrical Journal*, **48**, 876–886.

Bester, C.A., Hansen, C. (2009). A penalty function approach to bias reduction in non-linear panel models with fixed effects. *Journal of Business and Economic Statistics*, **27**, 131–148.

Brazzale, A.R., Davison, A.C., Reid, N. (2007). *Applied Asymptotics*. Cambridge University Press, Cambridge.

Carro, J.M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, **140**, 503–528.

Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society* B, **49**, 1–39.

Cruddas, A.M., Reid, N., Cox, D.R. (1989). A time series illustration of approximate conditional likelihood. *Biometrika*, **76**, 231–237.

DiCiccio, T.J., Martin, M.A., Stern, S.E., Young, G.A. (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society* B, **58**, 189–203.

DiCiccio, T.J., Stern, S.E. (1993). An adjustment to profile likelihood based on observed information. Technical Report, Department of Statistics, Stanford University.

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

Fernández-Val, I., Vella, F. (2011). Bias correction for two-step fixed effects panel data estimators. *Journal of Econometrics*, **163**, 144–162.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, **7**, 98–119.

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

Hajivassiliou, V. (2000). Some practical issues in maximum simulated likelihood. In: Mariano, R., Schuermann, T., Weeks, M.J. (Eds.) *Simulation-based Inference in Econometrics*, p. 71–99. Cambridge University Press, Cambridge.

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

Lancaster, T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies*, **69**, 647–666.

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

Pace, L., Salvan, A. (2006). A new motivation for adjustments of the profile likelihood. *Journal of Statistical Planning and Inference*, **136**, 3554–3564.

Pierce, D.A., Bellio, R. (2006). Effects of the reference set on frequentist inferences. *Biometrika*, **93**, 425–438.

Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, **90**, 533–549.

Severini, T. (1998). An approximation to the modified profile likelihood function. *Biometrika*, **85**, 403–411.

Severini, T. (1999). An empirical adjustment to the likelihood ratio statistic. *Biometrika*, **86**, 235–247.

Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.

Severini, T. (2002). Modified estimating functions. *Biometrika*, **89**, 333–343.

Wang, M., Hanfelt, J. (2008). Robust modified profile estimating function with application to the generalized estimating equation. *Journal of Statistical Planning and Inference*, **138**, 2029–2044.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	4	ML	-0.117	-0.118	0.058	0.130	0.118	0.847	0.362
		MPL	-0.005	-0.007	0.068	0.069	0.046	0.980	0.943
		IL	0.000	0.000	0.049	0.049	0.034	0.993	0.949
		Δ	0.957	0.941	-	0.753	0.857	-	-
	8	ML	-0.064	-0.065	0.042	0.076	0.065	0.932	0.599
		MPL	-0.001	-0.003	0.046	0.046	0.031	0.999	0.948
		IL	0.000	0.000	0.035	0.035	0.023	1.004	0.951
		Δ	0.984	0.954	-	0.732	0.810	-	-
500	4	ML	-0.122	-0.123	0.039	0.128	0.123	0.881	0.095
		MPL	-0.006	-0.006	0.047	0.047	0.032	1.016	0.949
		IL	-0.001	-0.001	0.034	0.034	0.023	1.018	0.956
		Δ	0.959	0.959	-	0.862	0.910	-	-
	8	ML	-0.064	-0.064	0.030	0.070	0.064	0.932	0.369
		MPL	-0.001	-0.002	0.032	0.032	0.022	1.001	0.951
		IL	0.000	0.000	0.024	0.024	0.017	1.004	0.953
		Δ	0.984	0.969	-	0.826	0.894	-	-
1000	4	ML	-0.121	-0.121	0.029	0.124	0.121	0.860	0.008
		MPL	-0.005	-0.006	0.034	0.035	0.023	0.999	0.946
		IL	0.000	0.000	0.025	0.025	0.017	1.007	0.955
		Δ	0.957	0.950	-	0.899	0.942	-	-
	8	ML	-0.063	-0.063	0.020	0.066	0.063	0.936	0.116
		MPL	-0.002	-0.002	0.022	0.022	0.015	1.008	0.953
		IL	0.000	0.000	0.017	0.017	0.012	1.005	0.953
		Δ	0.968	0.968	-	0.898	0.941	-	-

Table 1: Inference on β for the truncated regression model. Simulation results with 5,000 replications, $\beta = 1$.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	4	ML	-0.204	-0.204	0.042	0.208	0.204	0.847	0.001
		MPL	-0.008	-0.009	0.045	0.046	0.031	0.980	0.940
		IL	-0.001	-0.001	0.022	0.022	0.015	0.993	0.949
		Δ	0.966	0.961	-	0.871	0.915	-	-
	8	ML	-0.097	-0.097	0.029	0.101	0.097	0.932	0.063
		MPL	-0.003	-0.003	0.030	0.030	0.020	0.999	0.945
		IL	-0.001	-0.001	0.015	0.015	0.010	1.004	0.952
		Δ	0.979	0.979	-	0.826	0.885	-	-
500	4	ML	-0.205	-0.205	0.029	0.207	0.205	0.881	0.000
		MPL	-0.007	-0.008	0.032	0.033	0.023	1.016	0.941
		IL	-0.001	0.000	0.015	0.015	0.010	1.018	0.955
		Δ	0.971	0.961	-	0.906	0.882	-	-
	8	ML	-0.097	-0.097	0.021	0.099	0.097	0.932	0.002
		MPL	-0.002	-0.002	0.022	0.022	0.015	1.001	0.944
		IL	0.000	0.000	0.011	0.011	0.007	1.004	0.948
		Δ	0.979	0.979	-	0.875	0.911	-	-
1000	4	ML	-0.205	-0.205	0.021	0.206	0.205	0.860	0.000
		MPL	-0.006	-0.006	0.023	0.024	0.016	0.999	0.933
		IL	0.000	0.000	0.011	0.011	0.007	1.007	0.948
		Δ	0.971	0.971	-	0.933	0.955	-	-
	8	ML	-0.097	-0.097	0.014	0.098	0.097	0.936	0.000
		MPL	-0.003	-0.002	0.015	0.015	0.010	1.008	0.948
		IL	0.000	0.000	0.008	0.008	0.005	1.005	0.952
		Δ	0.969	0.979	-	0.922	0.946	-	-

Table 2: Inference on $\omega = \log \sigma$ for the truncated regression model. Simulation results with 5,000 replications, $\sigma = 1$.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	4	ML	0.759	0.729	0.265	0.804	0.729	0.699	0.012
		MML	-0.045	-0.047	0.075	0.087	0.063	1.459	0.980
		BC	-1.321	-1.182	0.570	1.438	1.182	0.157	0.000
		MPL	-0.032	-0.034	0.079	0.085	0.061	1.230	0.969
		IL	0.009	0.005	0.089	0.089	0.060	0.987	0.950
		Δ	0.969	0.960	-	1.006	0.999	-	-
	8	ML	0.249	0.246	0.083	0.263	0.246	0.857	0.063
		MML	0.015	0.013	0.059	0.061	0.040	1.019	0.952
		BC	-0.015	-0.016	0.054	0.056	0.038	1.086	0.958
		MPL	0.016	0.014	0.059	0.061	0.040	1.016	0.951
		IL	0.004	0.003	0.054	0.054	0.036	1.002	0.949
		Δ	0.951	0.955	-	0.967	0.981	-	-
500	4	ML	0.791	0.777	0.189	0.813	0.777	0.708	0.000
		MML	-0.063	-0.063	0.049	0.080	0.064	1.522	0.941
		BC	-1.413	-1.337	0.401	1.469	1.337	0.150	0.000
		MPL	-0.056	-0.057	0.051	0.076	0.058	1.308	0.911
		IL	0.005	0.002	0.062	0.063	0.042	0.995	0.949
		Δ	0.935	0.929	-	0.983	0.978	-	-
	8	ML	0.263	0.260	0.060	0.270	0.260	0.836	0.001
		MML	0.018	0.016	0.042	0.045	0.030	0.998	0.936
		BC	-0.016	-0.016	0.038	0.041	0.029	1.067	0.943
		MPL	0.019	0.017	0.042	0.046	0.031	1.002	0.938
		IL	0.003	0.002	0.038	0.038	0.026	0.993	0.948
		Δ	0.938	0.942	-	0.966	0.979	-	-
1000	4	ML	0.764	0.759	0.130	0.775	0.759	0.721	0.000
		MML	-0.067	-0.067	0.035	0.076	0.067	1.532	0.842
		BC	-1.314	-1.277	0.250	1.337	1.277	0.157	0.000
		MPL	-0.061	-0.062	0.036	0.071	0.062	1.315	0.793
		IL	0.002	0.001	0.043	0.043	0.029	1.010	0.950
		Δ	0.923	0.920	-	0.962	0.955	-	-
	8	ML	0.253	0.253	0.042	0.257	0.253	0.852	0.000
		MML	0.017	0.016	0.029	0.034	0.023	1.011	0.922
		BC	-0.011	-0.011	0.027	0.030	0.020	1.065	0.941
		MPL	0.017	0.017	0.029	0.034	0.023	1.012	0.920
		IL	0.001	0.001	0.027	0.027	0.018	0.995	0.950
		Δ	0.937	0.937	-	0.970	0.979	-	-

Table 3: Inference on β for the dynamic logit model. Simulation results with 5,000 replications, $\beta = 1$.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	4	ML	-2.688	-2.653	0.582	2.750	2.653	0.813	0.000
		MML	-0.574	-0.575	0.224	0.616	0.575	1.437	0.589
		BC	0.968	0.797	0.696	1.192	0.797	0.414	0.231
		MPL	-0.208	-0.209	0.211	0.296	0.218	1.438	0.964
		IL	0.001	-0.001	0.146	0.146	0.099	1.002	0.955
		Δ	0.923	0.922	-	0.942	0.953	-	-
	8	ML	-0.713	-0.714	0.178	0.735	0.714	0.944	0.013
		MML	-0.100	-0.101	0.151	0.181	0.128	1.026	0.911
		BC	-0.054	-0.056	0.143	0.153	0.104	1.075	0.953
		MPL	-0.085	-0.086	0.154	0.175	0.122	1.018	0.925
		IL	-0.002	-0.002	0.091	0.091	0.061	0.999	0.949
		Δ	0.883	0.882	-	0.870	0.907	-	-
500	4	ML	-2.481	-2.462	0.398	2.512	2.462	0.815	0.000
		MML	-0.540	-0.538	0.158	0.563	0.538	1.423	0.263
		BC	0.863	0.774	0.410	0.955	0.774	0.484	0.041
		MPL	-0.201	-0.202	0.150	0.251	0.203	1.417	0.928
		IL	0.001	0.000	0.107	0.107	0.072	0.993	0.948
		Δ	0.919	0.918	-	0.940	0.945	-	-
	8	ML	-0.712	-0.712	0.124	0.722	0.712	0.952	0.000
		MML	-0.089	-0.089	0.105	0.138	0.099	1.031	0.874
		BC	-0.045	-0.045	0.099	0.109	0.074	1.084	0.950
		MPL	-0.071	-0.072	0.106	0.128	0.089	1.029	0.907
		IL	0.000	-0.001	0.064	0.064	0.043	1.011	0.954
		Δ	0.900	0.899	-	0.903	0.931	-	-
1000	4	ML	-2.457	-2.446	0.284	2.474	2.446	0.807	0.000
		MML	-0.550	-0.549	0.116	0.562	0.549	1.376	0.017
		BC	0.757	0.718	0.251	0.798	0.718	0.547	0.002
		MPL	-0.239	-0.240	0.110	0.263	0.240	1.387	0.706
		IL	-0.001	0.001	0.076	0.076	0.052	0.989	0.947
		Δ	0.903	0.902	-	0.922	0.921	-	-
	8	ML	-0.725	-0.725	0.088	0.730	0.725	0.946	0.000
		MML	-0.093	-0.093	0.075	0.119	0.093	1.022	0.777
		BC	-0.047	-0.047	0.071	0.085	0.059	1.077	0.922
		MPL	-0.077	-0.076	0.076	0.108	0.081	1.017	0.841
		IL	0.000	0.000	0.045	0.045	0.030	0.998	0.948
		Δ	0.894	0.895	-	0.908	0.927	-	-

Table 4: Inference on ρ for the dynamic logit model. Simulation results with 5,000 replications, $\rho = 0.5$.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	4	ML	0.487	0.465	0.181	0.519	0.465	0.640	0.010
		MPL	-0.041	-0.042	0.042	0.059	0.045	1.404	0.944
		IL	0.007	0.005	0.054	0.054	0.036	0.995	0.950
		Δ	0.929	0.920	-	0.989	0.979	-	-
	8	ML	0.162	0.159	0.052	0.170	0.159	0.827	0.043
		MPL	0.007	0.005	0.035	0.035	0.023	1.029	0.953
		IL	0.002	0.001	0.032	0.032	0.021	0.990	0.946
		Δ	0.969	0.975	-	0.978	0.986	-	-
500	4	ML	0.507	0.497	0.125	0.523	0.497	0.665	0.000
		MPL	-0.062	-0.062	0.026	0.067	0.062	1.558	0.715
		IL	0.003	0.002	0.038	0.038	0.026	0.992	0.949
		Δ	0.883	0.879	-	0.940	0.924	-	-
	8	ML	0.167	0.167	0.037	0.171	0.167	0.821	0.000
		MPL	0.008	0.007	0.024	0.025	0.017	1.025	0.951
		IL	0.000	0.000	0.022	0.022	0.015	0.995	0.952
		Δ	0.952	0.958	-	0.980	0.987	-	-
1000	4	ML	0.483	0.478	0.084	0.490	0.478	0.686	0.000
		MPL	-0.064	-0.064	0.018	0.066	0.064	1.586	0.337
		IL	0.001	0.001	0.026	0.026	0.018	1.009	0.955
		Δ	0.869	0.868	-	0.914	0.900	-	-
	8	ML	0.163	0.162	0.025	0.165	0.162	0.839	0.000
		MPL	0.009	0.008	0.017	0.019	0.013	1.034	0.936
		IL	0.000	0.000	0.015	0.015	0.010	1.020	0.960
		Δ	0.945	0.951	-	0.973	0.980	-	-

Table 5: Inference on β for the dynamic probit model. Simulation results with 5,000 replications, $\beta = 1/1.6$.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	4	ML	-1.556	-1.531	0.387	1.603	1.531	0.779	0.001
		MPL	-0.137	-0.137	0.137	0.194	0.143	1.490	0.971
		IL	-0.001	0.000	0.092	0.092	0.063	1.004	0.953
		Δ	0.913	0.911	-	0.932	0.946	-	-
	8	ML	-0.424	-0.425	0.115	0.439	0.425	0.927	0.029
		MPL	-0.058	-0.058	0.097	0.113	0.077	1.009	0.916
		IL	-0.001	-0.001	0.057	0.057	0.039	0.987	0.947
		Δ	0.865	0.866	-	0.853	0.902	-	-
500	4	ML	-1.424	-1.417	0.263	1.448	1.417	0.785	0.000
		MPL	-0.128	-0.128	0.095	0.159	0.129	1.470	0.939
		IL	0.001	0.002	0.067	0.067	0.045	1.008	0.954
		Δ	0.911	0.911	-	0.933	0.939	-	-
	8	ML	-0.424	-0.424	0.078	0.431	0.424	0.948	0.001
		MPL	-0.052	-0.053	0.066	0.084	0.059	1.031	0.894
		IL	0.001	0.001	0.040	0.040	0.027	1.008	0.952
		Δ	0.879	0.877	-	0.887	0.919	-	-
1000	4	ML	-1.415	-1.409	0.183	1.427	1.409	0.806	0.000
		MPL	-0.160	-0.160	0.067	0.173	0.160	1.503	0.713
		IL	0.000	0.001	0.048	0.048	0.032	0.994	0.946
		Δ	0.887	0.887	-	0.909	0.907	-	-
	8	ML	-0.433	-0.433	0.057	0.437	0.433	0.926	0.000
		MPL	-0.055	-0.055	0.049	0.074	0.057	1.001	0.796
		IL	0.000	0.001	0.028	0.028	0.018	1.009	0.954
		Δ	0.873	0.875	-	0.888	0.906	-	-

Table 6: Inference on ρ for the dynamic probit model. Simulation results with 5,000 replications, $\rho = 0.5/1.6$.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	11	ML	0.158	0.156	0.060	0.169	0.156	0.883	0.164
		MPL	0.010	0.008	0.049	0.050	0.033	0.975	0.943
		MML	0.010	0.008	0.049	0.049	0.033	0.976	0.942
		BC	0.003	0.001	0.048	0.048	0.032	0.984	0.948
	16	ML	0.099	0.098	0.042	0.107	0.098	0.942	0.649
		MPL	0.005	0.004	0.037	0.037	0.025	1.000	0.950
		MML	0.005	0.004	0.037	0.037	0.025	1.001	0.950
		BC	0.003	0.002	0.036	0.037	0.025	1.002	0.953
500	11	ML	0.161	0.161	0.041	0.166	0.161	0.907	0.008
		MPL	0.010	0.010	0.033	0.034	0.023	1.002	0.947
		MML	0.009	0.009	0.033	0.034	0.023	1.003	0.949
		BC	0.002	0.002	0.032	0.032	0.022	1.012	0.954
	16	ML	0.097	0.097	0.029	0.102	0.097	0.961	0.479
		MPL	0.005	0.005	0.025	0.026	0.017	1.019	0.951
		MML	0.005	0.005	0.025	0.026	0.017	1.019	0.951
		BC	0.003	0.003	0.025	0.026	0.017	1.021	0.953
1000	11	ML	0.157	0.156	0.030	0.160	0.156	0.890	0.000
		MPL	0.009	0.008	0.024	0.026	0.017	0.983	0.936
		MML	0.009	0.008	0.024	0.025	0.017	0.983	0.937
		BC	0.002	0.001	0.024	0.024	0.016	0.988	0.949
	16	ML	0.097	0.096	0.021	0.099	0.096	0.952	0.220
		MPL	0.004	0.003	0.018	0.018	0.012	1.009	0.947
		MML	0.004	0.003	0.018	0.018	0.012	1.010	0.948
		BC	0.002	0.002	0.018	0.018	0.012	1.011	0.951

Table 7: Inference on β for the dynamic logit model, $T = 11, 16$. Simulation results with 5,000 replications, $\beta = 1$.

N	T	Method	B	MB	SD	RMSE	MAE	SE/SD	0.95 CI
250	11	ML	-0.466	-0.467	0.130	0.484	0.467	0.970	0.048
		MPL	-0.035	-0.035	0.120	0.125	0.085	1.006	0.943
		MML	-0.040	-0.039	0.119	0.126	0.086	1.012	0.941
		BC	-0.026	-0.027	0.113	0.116	0.078	1.062	0.958
	16	ML	-0.287	-0.286	0.097	0.303	0.286	0.984	0.090
		MPL	-0.013	-0.012	0.093	0.093	0.062	1.005	0.951
		MML	-0.014	-0.013	0.092	0.094	0.062	1.007	0.953
		BC	-0.015	-0.014	0.089	0.090	0.060	1.046	0.958
500	11	ML	-0.466	-0.466	0.094	0.476	0.466	0.955	0.000
		MPL	-0.035	-0.035	0.086	0.093	0.063	0.992	0.934
		MML	-0.040	-0.040	0.086	0.094	0.064	0.997	0.928
		BC	-0.025	-0.025	0.082	0.085	0.058	1.046	0.954
	16	ML	-0.290	-0.291	0.068	0.298	0.291	0.987	0.004
		MPL	-0.013	-0.014	0.065	0.066	0.045	1.007	0.952
		MML	-0.014	-0.015	0.065	0.067	0.045	1.008	0.952
		BC	-0.014	-0.015	0.062	0.064	0.044	1.050	0.959
1000	11	ML	-0.464	-0.464	0.065	0.469	0.464	0.961	0.000
		MPL	-0.035	-0.034	0.060	0.069	0.047	0.996	0.912
		MML	-0.040	-0.039	0.060	0.072	0.050	1.004	0.899
		BC	-0.024	-0.023	0.057	0.062	0.041	1.050	0.939
	16	ML	-0.288	-0.289	0.049	0.292	0.289	0.976	0.000
		MPL	-0.014	-0.014	0.047	0.048	0.033	0.998	0.939
		MML	-0.015	-0.015	0.046	0.049	0.033	1.001	0.938
		BC	-0.014	-0.015	0.045	0.047	0.032	1.038	0.946

Table 8: Inference on ρ for the dynamic logit model, $T = 11, 16$. Simulation results with 5,000 replications, $\rho = 0.5$.