Università Politecnica delle Marche
Dipartimento di Scienze Economiche e Sociali

Dynamic panel probit: finite-sample performance of alternative random-effects estimators

Riccardo Lucchetti Claudia Pigini

QUADERNI DI RICERCA n. 426*

February 2018

(*) La numerazione progressiva continua dalla serie denominata "Quaderni di ricerca — Dipartimento di economia"

Abstract

Estimation of random-effects dynamic probit models for panel data entails the so-called “initial conditions problem”. We argue that the relative finite-sample performance of the two main competing solutions is driven by the magnitude of the individual unobserved heterogeneity and/or of the state dependence in the data. We investigate our conjecture by means of a comprehensive Monte Carlo experiment and offer useful indications for the practitioner.

JEL Class.: C23, C25
Keywords: Dynamic panel probit; panel data; Monte Carlo study

Indirizzo: Università Politecnica delle Marche

Dynamic panel probit: finite-sample performance of alternative random-effects estimators[‡]

Riccardo Lucchetti *Claudia Pigini*

1 Introduction

Dynamic probit models for longitudinal data are nowadays largely employed in applied research. The most basic version of these models, without explanatory variables, can be written as:

$$\begin{aligned} y_{it}^* &= \gamma y_{it-1} + \alpha_i + \varepsilon_{it} \\ y_{it} &= 1\{y_{it}^* \geq 0\} \quad \text{for} \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T. \end{aligned} \tag{1}$$

The above formulation includes both the lagged response variable, which captures the so-called *true* state dependence (the effect of a past event on the probability of its future occurrence) via the parameter γ and the individual permanent unobserved effect α_i (Heckman, 1981a).

As customary with panel data models, estimation of this model may follow a fixed-effects or a random-effects approach. The former allows the individual unobserved heterogeneity to be arbitrarily correlated with the model covariates and several alternative solutions have been proposed to correct the bias of the Maximum Likelihood estimator arising because of the incidental parameters problem (Neyman and Scott, 1948).[1] However, random-effects models are often preferred as they allow for the identification of the effects associated with time-invariant explanatory variables, which are frequently of interest in many microeconomic applications.

In a random-effects setting, the recursive nature of the model requires that the process is initialised for y_{i0}, giving rise to the so-called "initial conditions" problem. Moreover, α_i and ε_{it} are usually assumed independent Gaussian random variables, with variances σ_α^2 and 1, respectively; this assumption introduces the problem of sequentially factoring the log-likelihood,

[‡] *We thank an anonymous referee for the comments and suggestions made to this paper, with the usual disclaimer.*

[1] See, among others, Carro (2007), Fernández-Val (2009), Bartolucci and Nigro (2010), Dhaene and Jochmans (2015), Bartolucci, Bellio, Salvan, and Sartori (2016)

since the initial observation y_{i0} and the individual effect α_i are correlated by construction.

Two main approaches have been proposed to deal with this problem. The solution proposed by Heckman (1981b) (HS, Heckman's Solution, henceforth) is derived by approximating the density $f(y_{i0}|\alpha_i)$ via a linearised reduced form equation as

$$y_{i0}^* = \eta + \theta\alpha_i + \varepsilon_{i0}, \qquad y_{i0} = 1\{y_{i0}^* \geq 0\}, \tag{2}$$

where θ is a nuisance parameter. An alternative approach was proposed by Wooldridge (2005) (WS, Wooldridge's Solution, from here on) by reversing the conditioning and specifying a distribution for the individual unobserved heterogeneity conditional on the initial value of the dependent variable $f(\alpha_i|y_{i0})$,[2] which amounts to writing

$$\alpha_i = \delta y_{i0} + \xi_i, \quad \xi_i \sim N(0, \sigma_\alpha^2), \tag{3}$$

where δ is a nuisance parameter.

WS has the virtue of being considerably simpler than HS from a computational viewpoint; however, it is well known that, for small T, it produces a severely biased estimator of γ (Arulampalam and Stewart, 2009; Akay, 2012). In this paper, we will make this statement more precise: we show by simulation that, for a given T, the major factor which determines the relative performance of HS versus WS is the magnitude of the individual effects, as measured by the σ_α^2 parameter, especially for positive values of the state dependence parameter γ. We further show that a simple descriptive statistic, the between variance of y_{it}, may offer the practitioner valid guidance in this respect.

The paper is as follows: in Section 2 we explain the main conjecture; in Section 3 we illustrate the simulation design and discuss the main results; section 4 concludes.

2 The main conjecture

We conjecture that, in the presence of strong persistence, HS is preferable, especially for small values of T. Strong persistence in y_{it} can be the result of both large positive values of γ and/or large values of $V(\alpha_i) = \sigma_\alpha^2$. Especially in the latter case, the auxiliary model in (2) is likely to yield a much better initialisation to the log-likelihood. In $f(\alpha_i|y_{i0})$, the conditioning variable is

[2] The original specification proposed by Wooldridge (2005) for this auxiliary model also includes functions of strictly exogenous covariates. See also Rabe-Hesketh and Skrondal (2013) and Skrondal and Rabe-Hesketh (2014) for the choice of the functional form.

binary by definition, and therefore provides information at most about the sign of α_i, but says very little about its magnitude.[3]

This intuition can be made more precise: from equation (1) it is straightforward to build the 2×2 Markov transition matrix for y_{it}:

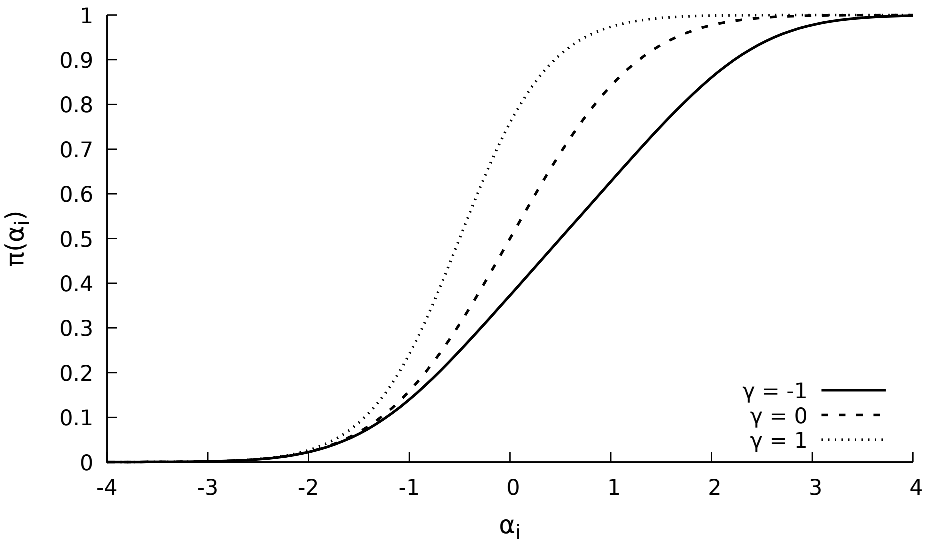
$$\Pi = \begin{bmatrix} 1 - \Phi(\alpha_i) & \Phi(\alpha_i) \\ 1 - \Phi(\alpha_i + \gamma) & \Phi(\alpha_i + \gamma) \end{bmatrix}$$

where $1 - \Phi(\alpha_i) = P(y_{it} = 0|y_{it-1} = 0)$, $\Phi(\alpha_i) = P(y_{it} = 1|y_{it-1} = 0)$ et cetera, and $\Phi(\cdot)$ is the cumulative Normal distribution; the steady-state probability for $y_{it} = 1$ is

$$\pi(\alpha_i) = \frac{\Phi(\alpha_i)}{\Phi(\alpha_i) + 1 - \Phi(\alpha_i + \gamma)}.$$

It is worth noting that the above expression is an increasing, invertible function of α_i (see Figure 1), so in principle the idea of inferring α_i given $\pi(\alpha_i)$ is sound. However, the function is practically flat for large α_i, especially when $\gamma = 1$.

Figure 1: $\pi(\alpha_i)$ as a function of α_i



A simple delta-method argument suggests that any estimate of $\pi(\alpha_i)$ would translate into a rather imprecise estimate of the individual effect α_i when its absolute value is large; especially so, for $\gamma \gg 0$; note, for example,

[3] If this conjecture is correct, WS for other models, such as the Poisson or the Tobit model, should exhibit a much better performance.

that for $\gamma = 1$, positive values of α_i are virtually observationally equivalent. In these cases, initialising the log-likelihood via $f(\alpha_i|y_{i0})$, as per WS, is likely to result in large finite-sample bias and/or variance.

In order to choose between the two approaches, it would be useful for the practitioner to have available a simple descriptive statistic flagging the cases in which σ_α and γ are likely to assume large positive values, so as to have an indication of when to prefer HS to WS. We argue that the *between* variance of $y_{i,t}$

$$V_b = \frac{1}{N}\sum_{i=1}^{N}(\bar{y}_i - \bar{y})^2, \tag{4}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^{N}\bar{y}_i$, is informative in this respect, in that V_b is increasing in the number of subjects in the sample exhibiting a high degree of persistence, which results in response configurations with almost all ones for some subjects, almost all zeros for others. This situation is likely to occur in presence of a large positive (negative) individual effect α_i, which increases the probability of y_{it} being equal to one (zero) in every time occasion, and a large positive value of γ, which reduces the likelihood of transitions between consecutive periods.[4] Moreover, since the dependent variable is binary, V_b lies by construction in the $[0, 0.25]$ interval, which makes it easy to interpret its magnitude.

3 Monte Carlo study

We perform a simulation study where each sample is generated by the model described in expression (1), with $T = 4, 8$ and $N = 500, 1000$. We generate 200 artificial datasets for each point in a bivariate grid of parameters $\gamma = -1, -0.9, \ldots, 0.9, 1$ and $\sigma_\alpha = 0.3, 0.4, \ldots, 1.4$; these are parameter values that we consider most likely to be relevant in real-world applications. For each subject i, we burn in 32 time periods, staring from an initial observation generated as $y_{i0} = 1\{\alpha_i + \varepsilon_{i0} > 0\}$. Estimation is carried out via the DPB `gretl` package (Lucchetti and Pigini, 2017).

In order to compare the relative performance of the Maximum Likelihood estimators of γ for the two alternative solutions, we use the following index:

$$\rho(\gamma, \sigma_\alpha) = \log\left[\frac{RMSE_W(\gamma, \sigma_\alpha)}{RMSE_H(\gamma, \sigma_\alpha)}\right],$$

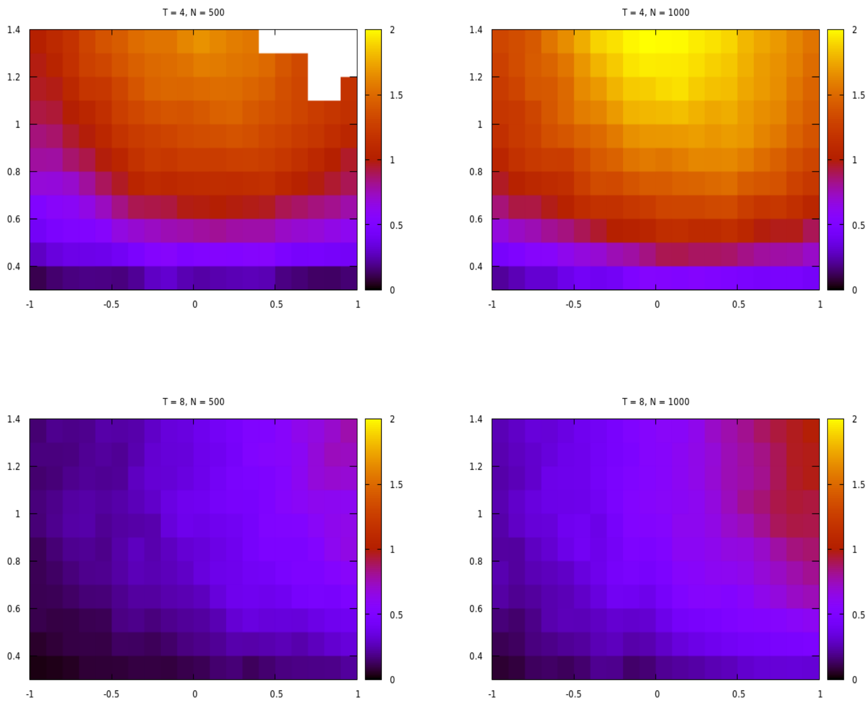
where $RMSE_W$ and $RMSE_H$ denote the root mean square error of $\hat{\gamma}$ from WS and HS, respectively, and the function arguments γ, σ_α link the index to

[4]Clearly, in a real-world situation V_b is also influenced by the distribution of observable covariates; therefore, in this case, it would be reasonable to use the variance of the residuals of a "between" OLS regression, instead.

each scenario identified by the bivariate grid for the state dependence and unobserved heterogeneity parameters.

Figure 2 shows the results, where $\rho(\gamma, \sigma_\alpha)$ is mapped over the grid for γ and σ_α, represented by the x and the left y axes, respectively. $\rho(\gamma, \sigma_\alpha)$ is uniformly positive over all the configurations considered. The relative advantage of HS is higher with small T, since WS is likely to yield a large bias in $\hat{\gamma}$ in this case, as also shown by Akay (2012). HS's edge in relative performance also increases with σ_α, which seems to affect the values of $\rho(\gamma, \sigma_\alpha)$ more importantly than γ. For $T = 8$, the largest difference in the relative performance occurs for large positive values of γ, increasing with σ_α. These results confirm the conjecture we proposed in Section 2: HS outperforms WS when persistence in the dependent variable is high, which may be the result of large individual effects and/or large positive values of the state dependence parameter.

Figure 2: Relative performance index $\rho(\gamma, \sigma_\alpha)$.



In each scenario, the values of γ are reported the x-axis, the values of σ_α are reported in the left y-axis. The right y-axis colour-codes the values of $\rho(\gamma, \sigma_\alpha)$. The uncharted area for $N = 500, T = 4$ corresponds to cases in which the simulation could not be completed, as at least one of the Monte Carlo replications generated a sample with no within variation.

Table 1: Correlation coefficient between $\rho(\gamma, \sigma_\alpha)$ and V_b

	T=4	T=8
$N = 500$	0.797	0.857
$N = 1000$	0.808	0.872

For each scenario, we have also computed the between variance V_b in (4). Table 1 shows that V_b works quite well as a predictor of $\rho(\gamma, \sigma_\alpha)$. Therefore V_b is a useful descriptive statistic in assessing whether HS actually has a relative advantage over WS: the higher the value of V_b, the greater the advantage.

4 Summary and indications for the practitioner

The superior computational simplicity of WS has made it popular among practitioners. However, it is well known that its computational advantage over HS may be outweighed, for small T, by its inferior small-sample performance. We add to this result by showing that, for a given T, the relative performance of HS versus WS is importantly driven by the magnitude of the individual effects and of the state dependence. Specifically, WS works rather poorly when persistence in the dependent variable is high. The between variance of the dependent variable, possibly adjusted for exogenous explanatory variables, can be used to choose between the two solutions, since it shows good predictive power for their relative performance.

References

Akay, A. (2012): "Finite-sample comparison of alternative methods for estimating dynamic panel data models," *Journal of Applied Econometrics*, 27(7), 1189–1204.

Arulampalam, W., and M. B. Stewart (2009): "Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators," *Oxford bulletin of economics and statistics*, 71(5), 659–681.

Bartolucci, F., R. Bellio, A. Salvan, and N. Sartori (2016): "Modified profile likelihood for fixed-effects panel data models," *Econometric Reviews*, 35(7), 1271–1289.

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(2), 719–733.

Carro, J. M. (2007): “Estimating dynamic panel data discrete choice models with fixed effects,” *Journal of Econometrics*, 140(2), 503–528.

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

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

Heckman, J. J. (1981a): “Heterogeneity and State Dependence,” *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press: Cambridge MA, Manski CF, McFadden (eds).

——— (1981b): “The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discreted data stochastic process,” *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press: Cambridge MA, Manski CF, McFadden (eds).

Lucchetti, R., and C. Pigini (2017): “DPB: Dynamic Panel Binary Data Models in gretl,” *Journal of Statistical Software, Articles*, 79(8), 1–33.

Neyman, J., and E. L. Scott (1948): “Consistent estimates based on partially consistent observations,” *Econometrica: Journal of the Econometric Society*, pp. 1–32.

Rabe-Hesketh, S., and A. Skrondal (2013): “Avoiding biased versions of Wooldridge’s simple solution to the initial conditions problem,” *Economics Letters*, 120(2), 346–349.

Skrondal, A., and S. Rabe-Hesketh (2014): “Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(2), 211–237.

Wooldridge, J. M. (2005): “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity,” *Journal of applied econometrics*, 20(1), 39–54.