Economics Letters ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Accounting for persistence in panel count data models. An application to the number of patents awarded

Stefanos Dimitrakopoulos

Department of Accounting, Finance and Economics, Oxford Brookes University, Oxford, OX331 HX, UK

HIGHLIGHTS

- We proposed a Poisson model for capturing persistence in panel counts.
- The model controls for dynamics, latent heterogeneity and serially correlated errors.
- In our empirical analysis we used data on patents granted.
- All sources of persistence were present with the serial error correlation being the strongest.

ARTICLE INFO

Article history:
Received 24 May 2018
Received in revised form 19 July 2018
Accepted 3 August 2018
Available online xxxx

JEL classification:
C1
C5
C11
C13

Keywords:
Dynamics
Initial conditions
Latent heterogeneity
Markov Chain Monte Carlo
Panel count data
Serial correlation

ABSTRACT

We propose a Poisson regression model that controls for three potential sources of persistence in panel count data; dynamics, latent heterogeneity and serial correlation in the idiosyncratic errors. We also account for the initial conditions problem. For model estimation, we develop a Markov Chain Monte Carlo algorithm. The proposed methodology is illustrated by a real example on the number of patents granted.

1. Introduction

There is a vast econometrics literature on the analysis of count data (Winkelmann, 2008; Cameron and Trivedi, 2013). In this paper we propose a Poisson model that accounts for three potential sources of the persistent behaviour of counts across economic units; true state dependence, spurious state dependence and serial error correlation.

True state dependence is modelled through a lagged dependent variable that controls for dynamic effects, spurious state dependence is captured by a latent random variable (Heckman, 1981) that controls for unobserved heterogeneity, while serial correlation in the idiosyncratic errors is assumed to follow a first-order stationary autoregressive process. The resulting model specification is a dynamic panel Poisson model with latent heterogeneity and serially correlated errors.

We also account for an inherent problem in our model, that of the endogeneity of the initial count for each cross-sectional unit (initial conditions problem). The assumption of exogenous initial conditions produces biased and inconsistent estimates (Fotouhi, 2005). To tackle this problem we apply the approach of Wooldridge (2005) that attempts to model the relationship between the unobserved heterogeneity and initial values.

In the context of Poisson regression analysis of event counts, researchers have proposed dynamic Poisson models with unobserved heterogeneity (Crépon and Duguet, 1997; Blundell et al., 2002) in order to disentangle true and spurious state dependence. Yet, the issue of persistence (true state dependence, spurious state dependence, serial error correlation) as well as the initial values problem have not been properly addressed in panel counts. This paper aspires to fill this gap.

To estimate the parameters of the proposed model, we develop a Markov Chain Monte Carlo (MCMC) algorithm, the efficiency of which is evaluated with a simulation study. We also conduct model

E-mail address: sdimitrakopoulos@brookes.ac.uk.

https://doi.org/10.1016/j.econlet.2018.08.004
0165-1765/

comparison. Our methodology is illustrated with an empirical example on patenting.

The paper is organized as follows. In Section 2 we set up the proposed model and in Section 3 we describe the posterior analysis. The empirical results are presented in Section 4. Section 5 concludes. An Online Appendix accompanies this paper.

2. Econometric framework

Let y_{it} be the observed count outcome for individual $i = 1, \ldots, N$ at time $t = 1, \ldots, T$, that follows the Poisson distribution with conditional mean λ_{it}

$$f(y_{it}; \lambda_{it}) = \frac{\lambda_{it}^{y_{it}} \exp(-\lambda_{it})}{y_{it}!}. \quad (1)$$

For λ_{it} we assume the following exponential mean function

$$\lambda_{it} = exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \gamma y_{it-1} + \varphi_i + \epsilon_{it}), \quad (2)$$

where $\mathbf{x}_{it} = (x_{1,it}, \ldots, x_{k,it})'$ is a vector of exogenous covariates[1] that contains an intercept, φ_i denotes the individual-specific random effect that controls for spurious state dependence, whereas the coefficient on y_{it-1} measures the strength of true state dependence.

Since y_{it} is non-negative, a positive coefficient γ makes the model explosive as $\gamma y_{it-1} > 0$. To overcome this problem we replace y_{it-1} in (2) by its logarithm, $\ln y_{it-1}$, and then use a strictly positive transformation y^*_{it-1} of the y_{it-1} values, when $y_{it-1} = 0$. In particular, we rescale only the zero values of y_{it-1} to a constant c, that is, $y^*_{it-1} = max(y_{it-1}, c)$, $c \in (0, 1)$; see also (Zeger and Qaqish, 1988). Therefore, expression (2) is replaced by

$$\lambda_{it} = exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \gamma \ln y^*_{it-1} + \varphi_i + \epsilon_{it}). \quad (3)$$

For the idiosyncratic error terms ϵ_{it}, we assume the following first-order stationary ($|\rho| < 1$) autoregressive specification

$$\epsilon_{it} = \rho\epsilon_{it-1} + \upsilon_{it}, \quad -1 < \rho < 1, \quad \upsilon_{it} \overset{i.i.d}{\sim} N(0, \sigma^2_\upsilon). \quad (4)$$

The random variables υ_{it} are independent and identically normally distributed across all i and t with mean zero and variance σ^2_υ. We also assume that υ_{it} and φ_i are mutually independent.

To tackle the initial values problem we follow the approach of Wooldridge (2005) and model φ_i as follows

$$\varphi_i = h_1 \ln y^*_{i0} + \bar{\mathbf{x}}_i'\mathbf{h}_2 + u_i, \quad u_i \sim N(0, \sigma^2_u), \quad i = 1, \ldots, N. \quad (5)$$

As before, if the first available count in the sample for individual i, y_{i0}, is zero, it is rescaled to a constant c, that is, $y^*_{i0} = max(y_{i0}, c)$, $c \in (0, 1)$. Also, $\bar{\mathbf{x}}_i$ is the time average of $\mathbf{x}_{it}$ and u_i is a stochastic disturbance, which is assumed to be uncorrelated with y_{i0} and $\bar{\mathbf{x}}_i$. For identification reasons, time-constant regressors that maybe included in $\mathbf{x}_{it}$ should be excluded from $\bar{\mathbf{x}}_i$.

To conduct Bayesian analysis we impose priors over the parameters $(\boldsymbol{\delta}, \mathbf{h}, \rho, \sigma^2_\upsilon, \sigma^2_u)$,

$$p(\boldsymbol{\delta}) \propto 1, \mathbf{h} \sim \mathbf{N}_{k+1}(\widetilde{\mathbf{h}}, \widetilde{\mathbf{H}}),$$

$$\rho \sim N(\rho_0, \sigma^2_\rho)I_{(-1,1)}(\rho), \ \sigma_\upsilon^{-2} \sim \mathcal{G}(\frac{e_1}{2}, \frac{f_1}{2}), \sigma^2_u \sim\sim \mathcal{IG}(\frac{e_0}{2}, \frac{f_0}{2}),$$

where $\boldsymbol{\delta} = (\boldsymbol{\beta}', \gamma)'$, $\mathbf{h} = (h_1, \mathbf{h}_2)'$, $\mathcal{G}$ denotes the gamma distribution and $\mathcal{IG}$ denotes the inverse gamma distribution. The prior distribution for $\boldsymbol{\delta}$ is flat. A truncated normal is imposed on ρ.

[1] Addressing the issue of potential violation of the exogeneity assumption in the context of the proposed model is a changeling econometric task and thus is left for future research; see also (Biewen, 2009) for potential treatment.

Table 1
Empirical results for the competing Poisson models.

	model 1	model 2	model 3	model 4
constant	0.1249 (0.1072)	0.0632 (0.1153)	−0.1350 (0.2030)	0.0294 (0.0303)
$\ln y^*_{it-1}$	0.0936* (0.0325)	0.2448* (0.0248)		0.9311* (0.0082)
SS	−0.0173 (0.0689)	0.0264 (0.0657)	0.4325* (0.1218)	0.0312* (0.0120)
ln *SIZE*	−0.0369 (0.0291)	−0.0012 (0.0316)	0.2843* (0.0511)	0.0205* (0.0059)
$\ln R_0$	0.2998* (0.0697)	0.3504* (0.0637)	0.4205* (0.0588)	0.2427* (0.0487)
$\ln R_1$	−0.0720 (0.0706)	−0.0777 (0.0718)	−0.0380 (0.0701)	−0.1659* (0.0681)
$\ln R_2$	0.0396 (0.0641)	0.0670 (0.0661)	0.1157 (0.0660)	−0.0514 (0.0646)
$\ln R_3$	0.0096 (0.0624)	0.0090 (0.0608)	0.0373 (0.0597)	−0.0294 (0.0599)
$\ln R_4$	0.0281 (0.0579)	0.0151 (0.0541)	0.0142 (0.0538)	0.0062 (0.0540)
$\ln R_5$	−0.0183 (0.0503)	0.0285 (0.0443)	0.0488 (0.0421)	0.0337 (0.0361)
YEAR=1976	−0.0384 (0.0227)	−0.041* (0.0177)	−0.0457* (0.0179)	−0.0222 (0.0176)
YEAR=1977	−0.0327 (0.0273)	−0.0372* (0.0181)	−0.0501* (0.0182)	0.0059 (0.0177)
YEAR=1978	−0.1457* (0.0294)	−0.1611* (0.0192)	−0.1776* (0.0189)	−0.1129* (0.0182)
YEAR=1979	−0.2002* (0.0341)	−0.1774* (0.0213)	−0.2316* (0.0199)	−0.0453* (0.0185)
σ^2_u	0.1091* (0.0386)	0.1481* (0.0208)	0.9942* (0.0963)	
σ^2_υ	0.0355* (0.0037)			
ρ	0.8311* (0.0751)			
BIC	−1390.21	−1411.47	−1432.98	−1439.74
CV	0.2095	0.1748	0.1744	0.1612

*Significant based on the 95% highest posterior density interval.
Standard deviations in parentheses.

3. Posterior analysis

3.1. MCMC algorithm

To estimate the model parameters, we follow closely the paper by Chib and Jeliazkov (2006) and develop a similar MCMC algorithm that augments the parameter space (Tanner and Wong, 1987) to include the latent variables $\{\lambda^*_{it}\}_{i\geq 1, t\geq 1}$, where $\lambda^*_{it} = \mathbf{w}_{it}'\boldsymbol{\delta} + \varphi_i + \epsilon_{it}$ and $\mathbf{w}_{it}' = (\mathbf{x}_{it}', \ln y^*_{it-1})$.

The details of the estimation method are given in the Online Appendix, where we also conduct a Monte Carlo experiment.

3.2. Model comparison

For model comparison we compute the marginal likelihood (ML). There are many ways to do that. One popular numerical method is the method of Chib (1995) and (Chib and Jeliazkov, 2001); see, also, (Chib et al., 1998). In this paper we use the

Please cite this article in press as: Dimitrakopoulos S., Accounting for persistence in panel count data models. An application to the number of patents awarded. Economics Letters (2018), https://doi.org/10.1016/j.econlet.2018.08.004.