Economics Letters 149 (2016) 116–119

Contents lists available at ScienceDirect

Economics Letters

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

Score-driven dynamic patent count panel data models

Szabolcs Blazsek [a,*], Alvaro Escribano [b]

[a] School of Business, Universidad Francisco Marroquín, Calle Manuel F. Ayau, 01010, Ciudad de Guatemala, Guatemala

[b] Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903, Getafe (Madrid), Spain

HIGHLIGHTS

- We propose the use of Dynamic Conditional Score (DCS) count panel data models.
- We compare the static, finite distributed lag, exponential feedback and DCS models.
- We use panel data for United States firms for period 1979–2000.
- We use the Poisson quasi-maximum likelihood estimator with fixed effects.
- The empirical results suggest that DCS-QAR is the best specification.

ARTICLE INFO

Article history:
Received 23 July 2016
Received in revised form
4 September 2016
Accepted 20 October 2016
Available online 29 October 2016

JEL classification:
C33
C35
C51
C52
O3

Keywords:
Research and development
Patent count panel data
Dynamic conditional score
Quasi-maximum likelihood

ABSTRACT

In this paper, we propose the use of Dynamic Conditional Score (DCS) count panel data models. We compare the statistical performance of the static model with different dynamic models: finite distributed lag, exponential feedback and different DCS models. For DCS, we consider random walk or quasi-autoregressive dynamics. We use panel data for a large cross section of United States firms for period 1979–2000, and the Poisson quasi-maximum likelihood estimator with fixed effects. The empirical results suggest that DCS has the best statistical performance.

1. Econometric model

Gourieroux et al. (1984a,b) and Wooldridge (1997a, 2002) motivate the use of the Quasi-Maximum Likelihood Estimator (QMLE) for count panel data models. For QMLE, a pseudo Log-Likelihood (LL) objective function is maximized, for which the pseudo probability distribution is within the Linear Exponential Family (LEF). An example of LEF is the Poisson distribution. In this paper, we use Poisson QMLE for patent count panel data models, hence $n_{it}|\mathcal{F}_t \sim \text{Poisson}(\lambda_{it})$ is the pseudo distribution for the count variable n_{it}. For this distribution, (i) $E(n_{it}|\mathcal{F}_t) = \lambda_{it}$, (ii) the log of the conditional probability mass function is $\ln f(n_{it}|\mathcal{F}_t) = -\lambda_{it} + n_{it}\ln\lambda_{it} - \ln(n_{it}!)$, (iii) the conditional score of n_{it} with respect to λ_{it} is $\partial \ln f(n_{it}|\mathcal{F}_t)/\partial\lambda_{it} = n_{it}/\lambda_{it} - 1 = s_{it}$, and (iv) $(s_{i1}, \ldots, s_{iT})$ is a martingale difference sequence with respect to $\mathcal{F}_t$, under the correct specification of the conditional mean of n_{it}.

We suggest count panel data models for which the error term e_{it} is possibly serially correlated. We introduce serial correlation into e_{it} by the dynamic variable Ψ_{it} that is updated by the pseudo conditional score s_{it-1}. We name these models as Dynamic Conditional Score (DCS) (Harvey, 2013) count panel data models. In the body of literature, Davis et al. (2003, 2005) and Harvey (2013) suggest dynamic time-series models for Poisson dependent variables updated by s_{it-1}. We extend those works, since we use (i) panel data models with unobserved effects, (ii) more parsimonious autoregressive dynamics for the impact of conditional score, and

* Corresponding author.
E-mail addresses: sblazsek@ufm.edu (S. Blazsek), alvaroe@eco.uc3m.es (A. Escribano).

http://dx.doi.org/10.1016/j.econlet.2016.10.026
0165-1765/

Table 1
Parameter estimates.

	SM	FDL	EFM
c	−1.3469*** (0.1669)	−1.2204*** (0.1424)	−1.1096*** (0.0702)
ζ_1	0.0932*** (0.0164)	0.0801*** (0.0128)	0.0588*** (0.0061)
ζ_2	−0.0155*** (0.0031)	−0.0125*** (0.0029)	−0.0132*** (0.0012)
ζ_3	−0.0061(0.0082)	−0.0103 (0.0106)	−0.0177*** (0.0021)
κ_0	0.9094*** (0.1363)	0.5689*** (0.1102)	0.4562*** (0.0311)
κ_1	NA	0.1349** (0.0652)	NA
κ_2	NA	0.1330*** (0.0471)	NA
κ_3	NA	0.1073(0.0750)	NA
κ_4	NA	0.0064(0.0183)	NA
κ_5	NA	0.0062(0.0395)	NA
ν_0	0.0112*** (0.0039)	0.0127*** (0.0029)	0.0024*** (0.0007)
ν_1	NA	0.0069*** (0.0018)	NA
ν_2	NA	−0.0044(0.0046)	NA
ν_3	NA	−0.0103*** (0.0036)	NA
ν_4	NA	−0.0184*** (0.0058)	NA
ν_5	NA	0.0358*** (0.0102)	NA
ξ_0	0.0073(0.0061)	0.0077(0.0081)	0.0029** (0.0013)
ξ_1	NA	−0.0012(0.0054)	NA
ξ_2	NA	0.0008(0.0126)	NA
ξ_3	NA	0.0295** (0.0118)	NA
ξ_4	NA	0.0084(0.0069)	NA
ξ_5	NA	−0.0741*** (0.0279)	NA
δ_1	0.5505*** (0.0634)	0.5381*** (0.0640)	0.0719*** (0.0256)
δ_2	−0.2465* (0.1386)	−0.2831** (0.1156)	−0.0481(0.0381)
δ_3	−0.6274(0.4508)	−0.6247* (0.3409)	−0.0905 (0.0754)
δ_4	−0.5605(0.3751)	−0.3794(0.3652)	−0.1855*** (0.0544)
α_1	NA	NA	0.8621*** (0.0279)

Notes: Not Available (NA). Robust standard errors are in parentheses.
* Significance at the 10% level.
** Significance at the 5% level.
*** Significance at the 1% level.

(iii) robust Poisson QMLE for statistical inference. The DCS count panel data model is

$$n_{it} = \exp(X_{it}'\beta)\upsilon_i e_{it} = \exp(X_{it}'\beta)\upsilon_i h(\Psi_{it})\epsilon_{it} \tag{1}$$

for $i = 1, \ldots, N$ firms and $t = 1, \ldots, T$ years, where X_{it} is a vector of explanatory variables, υ_i represents unobserved effects, e_{it} is a possibly serially correlated error term, Ψ_{it} is possibly serially correlated with $E(\Psi_{it}) = 0$, and ϵ_{it} is i.i.d. with $E(\epsilon_{it}) = 1$. We use the filter

$$\Psi_{it} = \alpha_1 \Psi_{it-1} + \gamma_1 s_{it-1} \tag{2}$$

that for $\alpha_1 = 1$ we name as the Random Walk (RW) specification, and for $|\alpha_1| < 1$ we name as the Quasi-Autoregressive (QAR) specification (Harvey, 2013). We initialize Ψ_{it} by the parameter $\Psi_0 = \Psi_{i0}$ for $i = 1, \ldots, N$. For h we consider alternative specifications, for which $h(\Psi_{it}) > 0$ and $h[E(\Psi_{it})] = 1$, such as:

$$h(\Psi_{it}) = \exp(\Psi_{it}) \tag{3}$$

$$h(\Psi_{it}) = \tanh(\Psi_{it}) + 1 \tag{4}$$

where tanh is the hyperbolic tangent function. We also use $h(\Psi_{it}) = [1 - \exp(-\Psi_{it})]/[1 + \exp(-\Psi_{it})] + 1$ and $h(\Psi_{it}) = 2F(\Psi_{it})$ (F is the distribution function of any symmetric continuous random variable centered at zero), but results are identical to those of Eq. (4).

2. Statistical inference

We estimate the parameters of DCS count panel data models by using QMLE with fixed effects Wooldridge (1997a, 2002). We maintain the following assumptions:

(A1) (pre-sample data) Pre-sample data $(n_{it}, X_{it} : t = 1, \ldots, P)$ are available. Let $\mathcal{F}_P$ denote the information set generated by pre-sample data.

(A2) (fixed effects) Replace υ_i by $p_i(\mathcal{F}_P) > 0$, where $p_i(\mathcal{F}_P)$ includes averages of n_{it} and X_{it} that are computed for the pre-sample data period (Blundell et al., 2002).

(A3) (correct specification of the mean) $E(n_{it}|X_{it}, \Psi_{it}, \mathcal{F}_P) = \exp(X_{it}'\beta)p_i(\mathcal{F}_P)h(\Psi_{it})$.

(A4) (martingale difference sequence) $(s_{it} : t = 1, \ldots, T)$ is a martingale difference sequence with respect to $\mathcal{F}_t = (X_{it}, \Psi_{it}, \mathcal{F}_P)$.

(A5) (exogeneity) All variables in X_{it} are predetermined (Blundell et al., 2002) (alternatively, all variables in X_{it} satisfy the sequential moment restrictions; Chamberlain, 1992 and Wooldridge, 1997a,b, 2002).

The use of pre-sample averages for fixed effects is motivated by the work of Blundell et al. (2002). We use QMLE with fixed effects, since QMLE with random effects is not feasible for DCS count panel data models, due to the latent υ_i that appears within the conditional score. We estimate the parameters consistently by using the pooled Poisson QMLE method with $\lambda_{it} = \exp(X_{it}'\beta)p_i(\mathcal{F}_P)h(\Psi_{it})$, by solving the maximization problem:

$$\arg\max_{\Theta} LL(\Theta) = \arg\max_{\Theta} \sum_{i=1}^{N}\sum_{t=1}^{T} -\lambda_{it} + n_{it}\ln\lambda_{it} - \ln(n_{it}!). \tag{5}$$

For the pooled Poisson QMLE, the pseudo score is $s_{it} = n_{it}/[\exp(X_{it}'\beta)p_i(\mathcal{F}_P)h(\Psi_{it})] - 1$, and we use the asymptotic distribution of parameter estimates of Wooldridge (1997a, 2002).

3. Data

The United States (US) utility patent dataset (source: MicroPatent LLC) includes the US Patent and Trademark Office (USPTO) patent number, application date, publication date, USPTO patent number of cited patents, three-digit US technological class, and company name. We perform all data procedures according to the work of Hall et al. (2001). We use the number of successful patent applications n_{it} for each firm and year. We measure spillovers of knowledge among firms by the log of the number of citations made of past patents of other firms of the same industry IA_{it} and of other industries IE_{it}. We use inflation-corrected log R&D expenses r_{it} to measure R&D investment (source: Standard & Poor's Compustat data files). We had created a match file and crossed the patent and firm datasets. The dataset includes 488,149 patents with application dates for period 1979–2000 (22 years) of 4476 US firms ($N = 4476$). We divide the full data window into two subperiods. First, the pre-sample data window is from 1979 to 1983 ($P = 5$ years). Second, the in-sample data window is from 1984 to 2000 ($T = 17$ years). It is noteworthy that Blazsek and Escribano (2010, 2016) use the same dataset.

4. Empirical results

We estimate five alternative multiplicative patent count panel data models. The first model is the Static Model (SM) for patent counts. For this model, $h(\Psi_{it}) = 1$ and $X_{it}'\beta$ is

$$X_{it}'\beta = c + \zeta_1 t + \zeta_2(t \times r_{it}) + \zeta_3 r_{it}^2 + \kappa_0 r_{it} + \nu_0 r_{it}\mathrm{IA}_{it} + \xi_0 r_{it}\mathrm{IE}_{it} \tag{6}$$

where the explanatory variables are motivated by Blazsek and Escribano (2010, 2016). The second model is the Finite Distributed Lag (FDL) model (Hausman et al., 1984), using

$$X_{it}'\beta = c + \zeta_1 t + \zeta_2(t \times r_{it}) + \zeta_3 r_{it}^2 + \sum_{k=0}^{5}\kappa_k r_{it-k} + r_{it}\sum_{k=0}^{5}\nu_k \mathrm{IA}_{it-k} + r_{it}\sum_{k=0}^{5}\xi_k \mathrm{IE}_{it-k} \tag{7}$$