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.

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).

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

$$n_{it} = \exp(X'_{it}\beta)v_i e_{it} = \exp(X'_{it}\beta)v_i h(\Psi_{it})\epsilon_{it} \quad (1)$$

for $i = 1, \ldots, N$ firms and $t = 1, \ldots, T$ years, where X_{it} is a vector of explanatory variables, v_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

$$\Psi_{it} = \alpha_1\Psi_{it-1} + \gamma_1 s_{it-1} \quad (2)$$

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:

$$h(\Psi_{it}) = \exp(\Psi_{it}) \quad (3)$$
$$h(\Psi_{it}) = \tanh(\Psi_{it}) + 1 \quad (4)$$

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 v_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 v_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:

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

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

$$X'_{it}\beta = c + \zeta_1 t + \zeta_2(t \times r_{it}) + \zeta_3 r_{it}^2 + \kappa_0 r_{it} + v_0 r_{it}IA_{it}$$
$$+ \xi_0 r_{it}IE_{it} \quad (6)$$

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

$$X'_{it}\beta = c + \zeta_1 t + \zeta_2(t \times r_{it}) + \zeta_3 r_{it}^2 + \sum_{k=0}^{5} \kappa_k r_{it-k}$$

$$+ r_{it}\sum_{k=0}^{5} v_k IA_{it-k} + r_{it}\sum_{k=0}^{5} \xi_k IE_{it-k} \quad (7)$$

Table 2

Parameter estimates.

	DCS-RW	DCS-RW	DCS-QAR	DCS-QAR
$h(\Psi_{it})$ specification:				
	$\exp(\Psi_{it})$	$\tanh(\Psi_{it}) + 1$	$\exp(\Psi_{it})$	$\tanh(\Psi_{it}) + 1$
Parameter estimates:				
c	1.3054***(0.0595)	2.3353***(0.0651)	0.1688***(0.0639)	3.9610***(0.1061)
ζ_1	0.1021***(0.0060)	0.0817***(0.0065)	0.0784***(0.0052)	0.0337***(0.0065)
ζ_2	−0.0130***(0.0012)	−0.0118***(0.0015)	−0.0126***(0.0012)	−0.0124***(0.0015)
ζ_3	0.0088***(0.0030)	0.0282***(0.0028)	0.0143***(0.0031)	0.0292***(0.0033)
κ_0	0.2482***(0.0751)	0.1855**(0.0749)	0.2491***(0.0786)	0.1948***(0.0736)
ν_0	0.0069***(0.0015)	0.0068***(0.0013)	0.0087***(0.0020)	0.0082***(0.0017)
ξ_0	0.0039(0.0032)	0.0053*(0.0027)	0.0007(0.0041)	0.0034(0.0033)
δ_1	0.7479***(0.0254)	0.7899***(0.0243)	0.6458***(0.0272)	0.6410***(0.0244)
δ_2	−0.0643(0.0794)	−0.1175(0.0789)	−0.0370(0.0841)	−0.0338(0.0781)
δ_3	−0.0510(0.0902)	−0.2010**(0.0966)	−0.1989(0.1343)	−0.3489***(0.1259)
δ_4	−0.1500(0.1108)	−0.5698***(0.1104)	−0.1990(0.1306)	−0.5194***(0.1148)
α_1	1.0000	1.0000	0.9384***(0.0383)	0.9670***(0.0356)
γ_1	0.1384**(0.0612)	0.0899*(0.0513)	0.1807***(0.0668)	0.1108*(0.0678)
Ψ_0	−1.9238***(0.0458)	−1.7946***(0.0463)	−0.7171***(0.0496)	−2.5749***(0.0592)

Notes: Robust standard errors are in parentheses.
* Significance at the 10% level.
** Significance at the 5% level.
*** Significance at the 1% level.

Table 3

Model diagnostics.

	SM	FDL	EFM	DCS-RW	DCS-RW	DCS-QAR	DCS-QAR
$h(\Psi_{it})$ specification:							
	$h(\Psi_{it}) = 1$	$h(\Psi_{it}) = 1$	Eq. (8)	Eq. (3)	Eq. (4)	Eq. (3)	Eq. (4)
$X_{it}\beta$ specification:							
	Eq. (6)	Eq. (7)	Eq. (6)	Eq. (6)	Eq. (6)	Eq. (6)	Eq. (6)
AR(1) test for residuals, robust SOLS:							
ρ	0.1559	0.0013	0.0028	0.0119	0.0002	0.0136	0.0011
p-value ρ	0.1234	0.8230	0.1273	0.4297	0.2515	0.1643	0.4118
Arellano–Bond test for residuals, robust optimal SGMM:							
ρ	0.0832	0.0010	−0.0001	−0.0012	0.0000	−0.0001	−0.0009
p-value ρ	0.2198	0.2512	0.7531	0.1729	0.8396	0.5969	0.8260
OITS	2.5179	3.8134	5.0244	4.1074	2.8717	3.4674	2.7618
p-value OITS	0.6414	0.4319	0.2848	0.3917	0.5795	0.4828	0.5984
APE for cross-section and time-series:							
	4.0243	2.0612	0.7581	1.2168	2.0724	1.5513	2.1620
Model selection:							
Mean LL	−3.7580	−3.5804	−2.0583	−2.0816	−1.9556	−2.0296	**−1.9176**
Mean AIC	7.5163	7.1615	4.1170	4.1636	3.9116	4.0596	**3.8356**
Mean BIC	7.5164	7.1618	4.1171	4.1637	3.9118	4.0598	**3.8358**
Mean HQC	7.5163	7.1615	4.1170	4.1636	3.9116	4.0597	**3.8357**

Notes: Bold numbers indicate the best statistical performance.

and $h(\Psi_{it}) = 1$. The third model is the Exponential Feedback Model (EFM) (Wooldridge, 2005), for which $X'_{it}\beta$ is Eq. (6) and

$$h(\Psi_{it}) = \exp[\alpha_1 \mathbb{1}(n_{it-1} > 0) \ln(n_{it-1})]. \tag{8}$$

The fourth and fifth models are DCS-RW and DCS-QAR, respectively, for which $X'_{it}\beta$ is Eq. (6) and $h(\Psi_{it}) = \exp(\Psi_{it})$ or $h(\Psi_{it}) = \tanh(\Psi_{it}) + 1$. We also estimated DCS with time effects, for which time trend was replaced by time dummy parameters. We found that the likelihood-based performance of DCS with time effects was inferior to that of DCS with Eq. (6), and γ_1 was not significant for DCS with time effects. For all models, we use $p_i(\mathcal{F}_P) = \exp(\delta_1 \ln \overline{n}_i + \delta_2 \overline{r}_i + \delta_3 \overline{r_{it} \text{IA}_{it}} + \delta_4 \overline{r_{it} \text{IE}_{it}})$, where the averages are for pre-sample data (Blundell et al., 2002).

Parameter estimates for all models are shown in Tables 1 and 2. We find that γ_1 is significant, at least, at the 10% level for DCS (see Table 2). Diagnostic tests are shown in Table 3. We use two tests for the serial correlation of ϵ_{it}. First, we estimate the AR(1) model $\epsilon_{it} = c^* + \rho\epsilon_{it-1} + u^*_{it}$ by using robust System Ordinary Least Squares (SOLS) (Wooldridge, 2002). Second, we use the Arellano and Bond (1991) dynamic panel data model $\epsilon_{it} = c^* + \rho\epsilon_{it-1} + v^*_i + u^*_{it}$, and estimate the first-differenced model $\Delta\epsilon_{it} = \rho\Delta\epsilon_{it-1} + \Delta u^*_{it}$

by using robust optimal System Generalized Method of Moments (SGMM) (Wooldridge, 2002). For SGMM, we use $(\epsilon_{it-2}, \ldots, \epsilon_{it-6})$ as instrumental variables. Both SOLS and SGMM suggest that ρ is not significant at the 10% level for all models (see Table 3). For SGMM, the Over-Identification Test Statistics (OITS) (Wooldridge, 2002) suggest that instrumental variables are always exogenous (see Table 3). Average Partial Effects (APE) of r_{it} for cross-section and time-series dimensions are shown in Table 3, and the evolution of APE of r_{it} for the cross-section dimension is shown in Fig. 1. APE for DCS of Fig. 1 corresponds to h and $X_{it}\beta$ of Eqs. (4) and (6), respectively. APE is interpreted as the average increase in λ_{it} due to a 1% increase in R&D expenses. We compare the in-sample performance of models by using: (i) mean LL; (ii) mean Akaike Information Criterion (AIC); mean Bayesian Information Criterion (BIC); (iv) mean Hannan–Quinn Criterion (HQC) (Hamilton, 1994) (see Table 3). All these metrics suggest that DCS-QAR is the best specification.
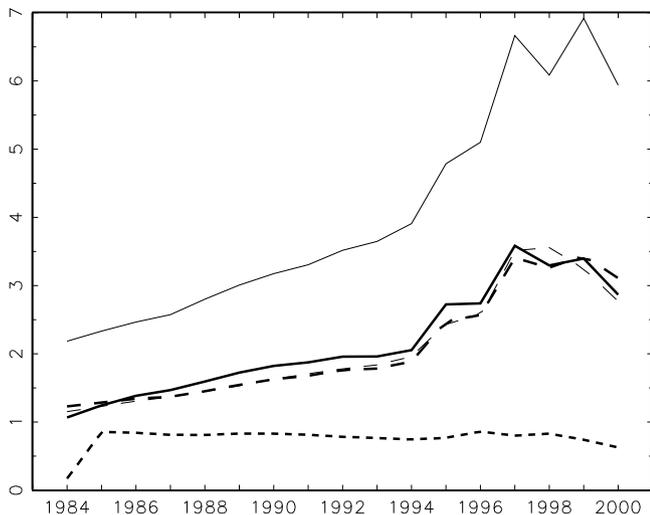
Fig. 1. Cross-section APE of r_{it}. *Notes*: SM (thin line); FDL (thin dashes); EFM (short thick dashes); RW (thick dashes); QAR (thick line).

Acknowledgments

References

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. 58, 277–297.

Blazsek, S., Escribano, A., 2010. Knowledge spillovers in US patents: A dynamic patent intensity model with secret common innovation factors. J. Econometrics 159, 14–32.

Blazsek, S., Escribano, A., 2016. Patent propensity, r&d and market competition: Dynamic spillovers of innovation leaders and followers. J. Econometrics 191, 145–163.

Blundell, R., Griffith, R., Windmeijer, F., 2002. Individual effects and dynamics in count data models. J. Econometrics 108, 113–131.

Chamberlain, G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.

Davis, R., Dunsmuir, W., Streett, S., 2003. Observation-driven models for Poisson counts. Biometrika 90, 777–790.

Davis, R., Dunsmuir, W., Streett, S., 2005. Maximum likelihood estimation for an observation driven model for Poisson counts. Methodol. Comput. Appl. 7, 149–159.

Gourieroux, C., Monfort, A., Trognon, A., 1984a. Pseudo maximum likelihood methods: Theory. Econometrica 52, 681–700.

Gourieroux, C., Monfort, A., Trognon, A., 1984b. Pseudo maximum likelihood methods: Applications to Poisson models. Econometrica 52, 701–720.

Hall, B., Jaffe, A.B., Trajtenberg, M., 2001. The NBER patent citation data file: Lessons, insights and methodological tools.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Harvey, A.C., 2013. Dynamic Models for Volatility and Heavy Tails. Cambridge University Press, Cambridge.

Hausman, J., Hall, B., Griliches, Z., 1984. Econometric models for count data with an application to the patents-r&d relationship. Econometrica 52, 909–938.

Wooldridge, J.M., 1997a. Quasi-likelihood methods for count data. In: Pesaran, M.H., Schmidt, P. (Eds.), Handbook of Applied Econometrics, Vol. 2. Blackwell, Oxford, pp. 352–406.

Wooldridge, J.M., 1997b. Multiplicative panel data models without the strict exogeneity assumption. Economet. Theor. 13, 667–678.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge.

Wooldridge, J.M., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. J. Appl. Econom. 20, 39–54.