UNIVERSIDAD CARLOS III DE MADRID working papers

UC3M Working papers
Economics
15-10
November, 2015
ISSN 2340-5031

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

DYNAMIC CONDITIONAL SCORE PATENT COUNT PANEL DATA MODELS

By Szabolcs Blazsek[a] and Alvaro Escribano[b,*]
[a]School of Business, Universidad Francisco Marroquín, Guatemala
[b]Department of Economics, Universidad Carlos III de Madrid, Spain

Abstract

We propose a new class of dynamic patent count panel data models that is based on dynamic conditional score (DCS) models. We estimate multiplicative and additive DCS models, MDCS and ADCS respectively, with quasi-ARMA (QARMA) dynamics, and compare them with the finite distributed lag, exponential feedback and linear feedback models. We use a large panel of 4,476 United States (US) firms for period 1979 to 2000. Related to the statistical inference, we discuss the advantages and disadvantages of alternative estimation methods: maximum likelihood estimator (MLE), pooled negative binomial quasi-MLE (QMLE) and generalized method of moments (GMM). For the count panel data models of this paper, the strict exogeneity of explanatory variables assumption of MLE fails and GMM is not feasible. However, interesting results are obtained for pooled negative binomial QMLE. The empirical evidence shows that the new class of MDCS models with QARMA dynamics outperforms all other models considered.

Keywords: patent count panel data models, dynamic conditional score models, quasi-ARMA model, research and development, patent applications.
JEL classification codes: C33, C35, C51, C52, O3.

* Corresponding author. Telefonica Chair of Economics of Telecommunications. Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903, Getafe (Madrid), Spain. Telephone: +34 916249854 (Escribano). E-mail addresses: sblazsek@ufm.edu (Blazsek); alvaroe@eco.uc3m.es (Escribano).

1. INTRODUCTION

PATENT ACTIVITY AND RESEARCH AND DEVELOPMENT (R&D) of firms were studied by the seminal works of Hausman, Hall, and Griliches (1984), Pakes (1985), and Jaffe (1986), who used panel data to investigate the relationships among input and output side measures of R&D activity, and various market value or accounting value based measures of firm performance. Other more recent works used dynamic patent count panel data models to study the relationships among patent counts, R&D and firm performance (Blundell, Griffith, and Windmeijer (2002); Wooldridge (2005); Blazsek and Escribano (2010), (2015); Bloom, Schankerman, and Van Reenen (2013)). In the present work we extend these models by suggesting a new class of dynamic patent count panel data models. By patent counts, we mean the number of successful patent applications of firms for a given year (Hausman, Hall, and Griliches (1984); Pakes (1985); Trajtenberg (1990); Lanjouw, Pakes, and Putnam (1998)). Although in recent patent databases the application date of patents is available with daily precision, this information is a noisy measure of the time of innovations. Therefore, following Hausman, Hall, and Griliches (1984), we aggregate patent counts over the year.

Harvey (2013, Chapter 5) noted that promising observation-driven time-series models of count data are the dynamic conditional score (DCS) Poisson models. In DCS Poisson models the dynamic equation is updated by the conditional score of the log-likelihood (LL) function, and the score is with respect to the time-varying conditional hazard parameter. The conditional score in these models discounts outliers and hence improves model fit. Harvey (2013) demonstrates the asymptotic theory of the maximum likelihood estimator (MLE) for different DCS models. These are location and scale DCS models for unrestricted and non-negative univariate continuous dependent variables, and multivariate DCS models for location, correlation and copula-based association. However, according to Harvey (2013, Section 5.11), the regularity conditions for the asymptotic distribution of MLE are not satisfied for DCS count data models based on the Poisson distribution.

In this paper, we propose new specifications of DCS models that allow us to estimate DCS

time-series models for firm-level panels of patent count data. These are the multiplicative DCS (MDCS) and additive DCS (ADCS) patent count panel data models. For both DCS models we consider quasi-ARMA (QARMA) dynamics (Harvey (2013)). In order to demonstrate the advantages of the new DCS count models, we compare them with the finite distributed lag (FDL) model (Hausman, Hall, and Griliches (1984)), exponential feedback model (EFM) (Wooldridge (2005)) and linear feedback model (LFM) (Blundell, Griffith, and Windmeijer (2002)). We estimate all models for the extended panel dataset of patent applications used by Blazsek and Escribano (2010, 2015). This panel dataset includes 4,476 US firms for period 1979 to 2000. Related to the statistical inference, we consider the advantages and disadvantages of alternative estimation methods: MLE, pooled negative binomial quasi-MLE (QMLE) and generalized method of moments (GMM). For the count panel data models of this paper the strict exogeneity assumption maintained in MLE fails and GMM is not feasible. Nevertheless, interesting empirical results are obtained by the pooled negative binomial QMLE for which the asymptotic distribution of parameter estimates is known (Wooldridge (1997a), (2002)). We test whether R&D expenditure was exogenous for each model, and we also compare the statistical performance of different models by the Pearson R-squared and deviance residual R-squared metrics (Cameron and Windmeijer (1996)). The results suggest that MDCS-QARMA is superior to FDL, EFM, LFM and ADCS-QARMA.

The remainder of this paper is organized as follows. Section 2 presents the firm-level panel dataset. Section 3 presents econometric modeling and statistical inference. Section 4 summarizes diagnostic tests and empirical results. Section 5 concludes.

2. DATA

Griliches (1990) states that the main advantages of patent data are the following: (i) by definition, patents are closely related to inventive activity; (ii) patent documents are objective because they are produced by an independent patent office and their standards change slowly over time; (iii) patent data are widely available in several countries over long periods of time and cover almost every field of innovation. Lanjouw and Schankerman (1999) and Hall, Jaffe,

and Trajtenberg (2001) also validate the use of patent statistics in economic research. We perform all data procedures according to the recommendations suggested by Hall, Jaffe, and Trajtenberg (2001). The source of the US utility patent dataset of this study is MicroPatent LLC. The US patent database includes the USPTO patent number, application date, publication date, USPTO patent number of cited patents, three-digit US technological class and company name (if the patent was assigned to a firm) for each patent. We use the application date to determine the time of an innovation, since inventors have incentive to apply for a patent as soon as possible after completing an innovation (Hall, Jaffe, and Trajtenberg (2001)). Company specific information is from Standard & Poor's (S&P) Compustat data files. For each firm, we use the book value for the sample midpoint year (Hausman, Hall, and Griliches (1984)) and R&D expenses for each year. We created a match file and crossed the patent dataset with the firm dataset via the six-digit Compustat CUSIP codes. Firm-specific data are corrected for inflation by using consumer price index (CPI) data (source: US Department of Labor, Bureau of Labor Statistics). The sample includes 488,149 US utility patents with application dates for period 1979 to 2000 ($T = 22$ years) of 4,476 US firms ($N = 4,476$). These data represent a case for which the cross-sectional dimension of the panel N is large, relative to its time-series dimension T. Therefore, we use the asymptotic theory presented by Wooldridge (1997a, 2002) for the estimation of count panel data models.

3. ECONOMETRIC MODELING AND STATISTICAL INFERENCE

3.1. *General Notation of Variables*

We observe a panel of patent application counts and other firm-specific variables of $i = 1, \ldots, N$ randomly selected firms for years $t = 1, \ldots, T$. n_{it} denotes the annual patent application count of firm i at period t; rd_{it} denotes the log of inflation adjusted R&D expenditure of firm i at period t; D_i takes the value one for firms in the drug, computer, scientific instrument, chemical and electronic components industries (high-tech industries), and zero otherwise; firm size b_i is the log of inflation adjusted book value for the sample midpoint year. For each count panel data

model we summarize the explanatory variables by the vector X_{it} for firm i and period t. For each model we present the components of X_{it} in the following section.

3.2. *Parameter Estimation by ML*

Hausman, Hall, and Griliches (1984) use the ML method to estimate the parameters of their patent count panel data models that include random effects (RE) or fixed effects (FE). We denote both RE and FE by α_i. MLE maintains the following assumptions:

(ML1) $n_{it}|(X_{i1}, \ldots, X_{iT}, \alpha_i) \sim$ Poisson(λ_{it}), where $\lambda_{it} = \lambda(X_{it}, \theta)$ is the conditional hazard parameter of the Poisson distribution. This implies strict exogeneity of all explanatory variables, conditional on α_i.

(ML2) λ_{it} is modeled by the exponential function: $\lambda_{it} = \lambda(X_{it}, \theta) = \exp(X_{it}\beta + \ln \alpha_i) = \exp(X_{it}\beta)\alpha_i$.

(ML3) $n_{it}|(X_{i1}, \ldots, X_{it}, \alpha_i)$ and $n_{is}|(X_{i1}, \ldots, X_{is}, \alpha_i)$ for $t \neq s$ are independent.

(ML4) α_i is independent and identically distributed (i.i.d.) with Gamma(1,δ) distribution. This implies that $E(\alpha_i) = 1$ and $\text{Var}(\alpha_i) = \delta$.

MLE is given by those parameters that maximize LL of the patent application count time series, i.e., $\hat{\theta} = \arg\max_\theta LL(n_{i1}, \ldots, n_{iT}; \theta)$. For RE MLE, (ML1) to (ML4) are maintained, thus α_i is assumed to be independent of X_{it}. For FE MLE, (ML1) to (ML3) are maintained, hence $\alpha_1, \ldots, \alpha_N$ are not treated as latent i.i.d. variables, but rather as constant parameters. As a consequence, the coefficients of time-constant explanatory variables are not identified and α_i is possibly not independent of X_{it}. If the assumptions maintained hold, then the ML method will provide an efficient estimator (Wooldridge (1997a)).

In the following, we present LL for RE MLE and FE MLE. First, for patent count data models with RE, the marginal density of $(n_{it}|X_{i1}, \ldots, X_{it})$ can be obtained by integrating out

α_i from the joint density of $(n_{it}, \alpha_i | X_{i1}, \ldots, X_{it})$, as follows:

$$(3.1)\quad f(n_{it}|X_{i1},\ldots,X_{it}) = \int_0^\infty \frac{\exp(-\lambda_{it})\lambda_{it}^{n_{it}}}{n_{it}!} \times \frac{\delta^\delta \alpha_i^{\delta-1}\exp(-\delta\alpha_i)}{\Gamma(\delta)} d\alpha_i,$$

where Γ is the gamma function. The integrand of this equation is the product of the conditional probability mass function of $n_{it}|(X_{i1},\ldots,X_{it},\alpha_i) \sim \text{Poisson}(\lambda_{it})$ and the marginal density function of $\alpha_i \sim \text{Gamma}(1,\delta)$. Under (ML2), equation (3.1) can be written as

$$(3.2)\quad f(n_{it}|X_{i1},\ldots,X_{it}) = \frac{\delta^\delta[\exp(X_{it}\beta)]^{n_{it}}}{n_{it}!\Gamma(\delta)} \int_0^\infty \exp\left\{-\alpha_i\left[\exp(X_{it}\beta)+\delta\right]\right\} \alpha_i^{n_{it}+\delta-1} d\alpha_i.$$

In order to evaluate the integral we use the formula

$$(3.3)\quad \int_0^\infty \exp(-ax)x^b dx = \frac{\Gamma(b+1)}{a^{b+1}}.$$

We substitute $x = \alpha_i$, $a = \exp(X_{it}\beta) + \delta$ and $b = n_{it} + \delta - 1$ into equation (3.3) and obtain

$$(3.4)\quad f(n_{it}|X_{i1},\ldots,X_{it}) = \frac{\delta^\delta[\exp(X_{it}\beta)]^{n_{it}}}{n_{it}!\Gamma(\delta)} \times \frac{\Gamma(n_{it}+\delta)}{[\exp(X_{it}\beta)+\delta]^{n_{it}+\delta}}.$$

Hence, LL for the exponential patent count data model with RE is

$$(3.5)\quad \begin{aligned} LL(X_{i1},\ldots,X_{iT},\theta) &= \textstyle\sum_{i=1}^N \sum_{t=1}^T l_{it}(\theta) = \sum_{i=1}^N \sum_{t=1}^T \delta\ln(\delta) + n_{it}(X_{it}\beta) \\ &\quad + \ln\Gamma(n_{it}+\delta) - \ln\Gamma(\delta) - (n_{it}+\delta)\ln[\exp(X_{it}\beta)+\delta]. \end{aligned}$$

From equation (3.5) we excluded $-\ln(n_{it}!)$, since it does not depend on the parameters. Second, for exponential patent count data models with FE the conditional hazard parameter is $\lambda_{it} = \exp(X_{it}\beta)\alpha_i$. This gives the following form of LL under (ML1):

$$(3.6)\quad \begin{aligned} LL(X_{i1},\ldots,X_{iT},\theta) &= \textstyle\sum_{i=1}^N \sum_{t=1}^T [n_{it}\ln\lambda_{it} - \ln(n_{it}!) - \lambda_{it}] \\ &= \textstyle\sum_{i=1}^N \sum_{t=1}^T [n_{it}(X_{it}\beta + \ln\alpha_i) - \ln(n_{it}!) - \exp(X_{it}\beta + \ln\alpha_i)]. \end{aligned}$$

We can approximate FE α_i by solving the first-order condition $\partial LL/\partial(\ln \alpha_i) = 0$ that gives

$$(3.7)\quad \alpha_i = \frac{\sum_{t=1}^{T} n_{it}}{\sum_{t=1}^{T} \exp(X_{it}\beta)}.$$

We substitute this result into equation (3.6) and introduce the notation

$$(3.8)\quad p_{it} = \exp(X_{it}\beta) / \left[\sum_{s=1}^{T} \exp(X_{is}\beta)\right].$$

Under this notation, LL for the exponential patent count data model with FE is

$$(3.9)\quad LL(X_{i1}, \ldots, X_{iT}, \theta) = \sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}(\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[n_{it} \ln\left(p_{it}\sum_{s=1}^{T} n_{is}\right) - p_{it}\sum_{s=1}^{T} n_{is}\right].$$

From equation (3.9) we excluded $-\ln(n_{it}!)$, since it does not depend on the parameters. Hausman, Hall, and Griliches (1984) note that this LL is conditional on the sum of the number of patents in the sample, i.e., $\sum_{s=1}^{T} n_{is}$. The asymptotic variance of $\hat{\theta}$ for both RE MLE and FE MLE is given by $A^{-1}BA^{-1}/N$. This is due to the following result shown by Wooldridge (1997a, 2002): $\sqrt{N}(\hat{\theta} - \theta) \rightarrow_d N(0, A^{-1}BA^{-1})$, where $A = E[-H_i(\theta)]$; $B = E[s_i(\beta)s_i(\beta)']$; $H_i(\theta) = \sum_{t=1}^{T} \nabla_\theta s_{it}(\theta)$; $s_i(\theta) = \sum_{t=1}^{T} s_{it}(\theta) = \sum_{t=1}^{T} \nabla_\theta l_{it}(\theta)$. Moreover, $s_{it}(\theta)$ denotes the score with respect to θ for observation n_{it}, and l_{it} for RE MLE and FE MLE is given by equations (3.5) and (3.9), respectively. Consistent estimators of A and B are given by sample averages.

In the following we present the reasons why MLE is probably not the most adequate estimation method for the count data models of this paper. First, for the FDL model k lags of R&D are considered, $X_{it} = (1, D_i, b_i, rd_{it}, \ldots, rd_{it-k})$. For EFM and LFM, X_{it} also includes the first lag of patent count and the initial condition, $X_{it} = (1, n_{it-1}, n_{i1}, D_i, b_i, rd_{it}, \ldots, rd_{it-k})$. For all DCS count panel data models, X_{it} includes several lags of the conditional score variable u_{it} instead of the first lag of patent count. For example, for MDCS-QMA(q), $X_{it} = (1, u_{it-1}, \ldots, u_{it-q}, n_{i1}, D_i, b_i, rd_{it}, \ldots, rd_{it-k})$. In all these models, the strict exogeneity assumption of (ML1) may fail. For the FDL model (ML1) will fail if patent count affects future R&D

expenditure. For EFM and LFM, (ML1) fails since the first lag of patent count n_{it-1} is included as explanatory variable. For DCS count data models (ML1) fails since the conditional score u_{it} is a transformation of the patent count n_{it}, hence DCS models also include lagged patent counts in the model specification. Second, MLE assumes that unobserved effects α_i are included in λ_{it}. For the DCS count panel data models considered in this paper, LL is not available in closed form when unobserved effects are included in λ_{it}. This is due to the fact that lags of λ_{it} appear in $X_{it}\beta$, within the conditional score terms. This motivates the application of pooled panel data models for which unobserved effects are not considered in the model formulation. As MLE may not be a consistent estimator of parameters for the count panel data models of this paper, therefore the pooled negative binomial QMLE (Wooldridge (1997a), (2002)) or GMM (Chamberlain (1992); Wooldridge (1997b)) estimators that do not require strict exogeneity of all explanatory variables, are possibly more adequate.

3.3. *Multiplicative Exponential Count Panel Data Models*

All models presented in this section are formulated for $E(n_{it}|X_{i1},\ldots,X_{it}) = \lambda_{it} = \exp(X_{it}\theta)$. For this functional form of the conditional mean, the explanatory variables and the error term are multiplicative. In this section, we review the pooled FDL model (Hausman, Hall, and Griliches (1984)) and its extension, the pooled EFM (Wooldridge (2005)). In these models we do not condition on unobserved effects α_i in the count data model, thus we consider pooled count panel data models. The pooled FDL model (Hausman, Hall, and Griliches (1984)) is

$$\lambda_{it} = \exp(X_{it}\theta) = \exp[\mu_0 + \gamma_1 t + \gamma_2(t \times rd_{it}) + \gamma_3 D_i + \gamma_4 b_i + \beta_5(L)rd_{it}], \tag{3.10}$$

where t is linear time trend; D_i indicates if the firm is in a high-tech industry; b_i measures firm size; $\beta_5(L) = \sum_{k=0}^{5}\beta_k L^k$ is the lag polynomial of five lags, which captures contemporaneous and lagged impact of R&D expenses rd_{it} on patent counts.

The pooled EFM (Wooldridge (2005)) includes the first lag of patent counts n_{it-1} in the

conditional hazard and controls the initial condition n_{i1} of the dynamic process, as follows:

$$(3.11)\quad \lambda_{it} = \exp(X_{it}\theta) = \exp[\mu_0 + \gamma_1 t + \gamma_2(t \times rd_{it}) + \gamma_3 D_i + \gamma_4 b_i + \gamma_5 n_{i1} + \beta_5(L)rd_{it} + \phi_1 n_{it-1}]$$

3.4. *Multiplicative DCS Count Panel Data Models*

We propose a new formulation for the conditional expectation of patent count. The MDCS count data model is formulated for $E(n_{it}|X_{i1}, \ldots, X_{it}) = \lambda_{it} = \exp(X_{it}\theta)$. This implies that MDCS is also a multiplicative count panel data model, and we do not condition on unobserved effects α_i in the conditional expectation of patent count. MDCS involves lags of the dynamic term u_{it}, which is defined as follows:

$$(3.12)\quad u_{it} = \frac{n_{it} - \lambda_{it}}{\lambda_{it}} = \frac{n_{it}}{\lambda_{it}} - 1.$$

The same innovation term is suggested by Davis, Dunsmuir, and Streett (2003, 2005) and Harvey (2013, Section 5.11), who propose the DCS Poisson model for time-series data and estimate it by the ML method. They define u_{it} according to equation (3.12), since it coincides with the DCS of the Poisson LL with respect to the conditional hazard parameter λ_{it}. This can be demonstrated as follows. The conditional probability mass function of the Poisson random variable with conditional expectation λ_{it} is

$$(3.13)\quad f(n_{it}|X_{i1}, \ldots, X_{it}) = \frac{\exp(-\lambda_{it})\lambda_{it}^{n_{it}}}{n_{it}!}.$$

The partial derivative of the log of this function with respect to λ_{it} is

$$(3.14)\quad \frac{\partial \ln f(n_{it}|X_{i1}, \ldots, X_{it})}{\partial \lambda_{it}} = \frac{n_{it}}{\lambda_{it}} - 1 = u_{it}.$$

In a time-series framework Davis, Dunsmuir, and Streett (2003, 2005) and Harvey (2013, Section 5.11) suggest using u_{it} as innovation term for the DCS Poisson model. Furthermore, Harvey

(2013) also suggests DCS time-series models with first-order autoregressive formulation. According to these a possible first-order count panel data model would be

$$(3.15)\quad \lambda_{it} = \lambda(X_{it}, \theta) = \exp(X_{it}\theta) = \exp(\phi_1 \ln \lambda_{it-1} + \theta_1 u_{it-1} + Y_{it}\tilde{\theta})$$

with $|\phi_1| < 1$ for covariance stationarity, $\theta = (\phi_1, \theta_1, \tilde{\theta})$ and

$$(3.16)\quad Y_{it}\tilde{\theta} = \mu_0 + \gamma_1 t + \gamma_2(t \times rd_{it}) + \gamma_3 D_i + \gamma_4 b_i + \gamma_5 n_{i1} + \beta_5(L) rd_{it}.$$

Harvey (2013, p. 37, Theorem 1) demonstrates the information matrix for general first-order DCS models. Unfortunately, the conditions of this theorem do not hold for the first-order DCS Poisson model of equations (3.15) and (3.16). For time-series data Davis, Dunsmuir, and Streett (2003, 2005) use an alternative specification for the conditional mean of n_{it} by considering several lags of u_{it}, but they do not include autoregressive terms in their model. These authors derive the information matrix for MLE and show that there exists an asymptotic distribution of MLE. Nevertheless, Harvey (2013, Section 5.11) notes that the central limit theorem is currently unavailable for MLE for this model. Based on Davis, Dunsmuir, and Streett (2003, 2005), the MDCS count panel data model is

$$(3.17)\quad \lambda_{it} = \lambda(X_{it}, \theta) = \exp(X_{it}\theta) = \exp(\Psi_{it} + Y_{it}\tilde{\theta})$$

$$(3.18)\quad \Psi_{it+1} = \theta_0 u_{it} + \theta_1 u_{it-1} + \ldots + \theta_q u_{it-q},$$

where $\theta = (\theta_0, \ldots, \theta_q, \tilde{\theta})$. Following the terminology of Harvey (2013, p. 63), we name this model MDCS-Quasi-MA(q) or MDCS-QMA(q). We also define a more compact formulation with infinite lags of u_{it} by the next MDCS-QAR(1) model:

$$(3.19)\quad \lambda_{it} = \lambda(X_{it}, \theta) = \exp(X_{it}\theta) = \exp(\Psi_{it} + Y_{it}\tilde{\theta})$$

(3.20) $\Psi_{it+1} = \phi_1 \Psi_{it} + \theta_0 u_{it},$

where $|\phi_1| < 1$ and $\theta = (\phi_1, \theta_0, \tilde{\theta})$. Furthermore, we combine the previous models to obtain the following MDCS-QARMA(p,q) model:

(3.21) $\lambda_{it} = \lambda(X_{it}, \theta) = \exp(X_{it}\theta) = \exp(\Psi_{it} + Y_{it}\tilde{\theta})$

(3.22) $\Psi_{it+1} = \phi_1 \Psi_{it} + \ldots + \phi_p \Psi_{it-p} + \theta_0 u_{it} + \theta_1 u_{it-1} + \ldots + \theta_q u_{it-q}.$

3.5. *Additive DCS Count Panel Data Models*

The previous count data models with exponential conditional mean function assume a multiplicative form of the conditional expectation of patent count. Nevertheless, there are several alternative formulations of the conditional mean of patent count in the literature. Examples are the Box-Cox-like model (Wooldridge (1997a)) and LFM (Blundell, Griffith, and Windmeijer (2002)). All additive count panel data models of this section are formulated for $E(n_{it}|X_{i1}, \ldots, X_{it}) = \lambda_{it}$. This implies that we do not condition on unobserved effects α_i in the conditional expectation of patent count. We start with the pooled LFM, which is specified as

(3.23) $\lambda_{it} = \lambda(X_{it}, \theta) = \phi_1 n_{it-1} + \exp(Y_{it}\tilde{\theta}),$

where $0 < \phi_1 < 1$ and $\theta = (\phi_1, \tilde{\theta})$. $Y_{it}\tilde{\theta}$ is defined in equation (3.16). In the following, we propose several ADCS specifications. The ADCS-QMA(q) count panel data model is

(3.24) $\lambda_{it} = \lambda(X_{it}, \theta) = \Psi_{it} + \exp(Y_{it}\tilde{\theta})$

(3.25) $\Psi_{it+1} = \theta^* + \theta_0 u_{it} + \theta_1 u_{it-1} + \ldots + \theta_q u_{it-q},$

where $\theta^* = \theta_0 + \ldots + \theta_q$ and $\theta = (\theta_0, \ldots, \theta_q, \tilde{\theta})$. We include θ^* in equation (3.25) to ensure the positivity of λ_{it}. A more compact form of this ADCS specification with less parameters, but

with infinite lags of u_{it}, is the following ADCS-QAR(1) model:

(3.26) $\Psi_{it+1} = \theta_0 + \phi_1 \Psi_{it} + \theta_0 u_{it}$,

where $|\phi_1| < 1$ and $\theta = (\phi_1, \theta_0, \tilde{\theta})$. We include θ_0 as constant parameter in equation (3.26) to ensure the positivity of λ_{it}. Finally, we also define a more general dynamic formulation, the ADCS-QARMA(p,q) model, as follows:

(3.27) $\Psi_{it+1} = \theta^* + \phi_1 \Psi_{it} + \ldots + \phi_p \Psi_{it-p} + \theta_0 u_{it} + \theta_1 u_{it-1} + \ldots + \theta_q u_{it-q}$,

where $\theta^* = \theta_0 + \ldots + \theta_q$ and $\theta = (\phi_1, \ldots, \phi_p, \theta_0, \ldots, \theta_q, \tilde{\theta})$.

3.6. *Parameter Estimation by QML*

In this section, we present the details of the pooled negative binomial QMLE method. We use the methodology presented by Wooldridge (1997a, Sections 4.2 and 9.2; 2002, Section 19.6). The main advantage of the pooled negative binomial QMLE estimator with respect to MLE is that it requires weaker assumptions for consistent estimation. First, MLE assumes a specific conditional distribution of patent count that is not needed for the pooled negative binomial QMLE. Second, MLE assumes strict exogeneity for all explanatory variables that may fail, for example, due to lagged dependent variables included as explanatory variables, or due to feedback effects of past patent counts on future R&D expenses. Nevertheless, the pooled negative binomial QMLE can consistently estimate models with lagged dependent variables or other variables that are not strictly exogenous explanatory variables (Wooldridge (1997a)). Third, the pooled negative binomial QMLE does not consider α_i in the model specification. In the count data models estimated by this method $E(n_{it}|X_{i1}, \ldots, X_{it}) = \lambda(X_{it}, \theta) = \lambda_{it}$, hence we do not condition on α_i. This is useful for both DCS count panel data models, where LL is not available in closed form due to the latent α_i term within $X_{it}\beta$.

The negative binomial QMLE can be related to the MLE of a count panel data model

with RE for the case when the conditional distribution of patent count is Poisson and α_i has gamma distribution. A well-known choice for the parameters of gamma distribution for α_i is Gamma$(1, \delta)$. By integrating out RE from the joint density of patent count and RE, we obtain a negative binomial probability specification of the second kind that coincides with the objective function of the negative binomial QMLE (Hausman, Hall, and Griliches (1984); Cameron and Trivedi (1986); Wooldridge (1997a)). Furthermore, the use of the log negative binomial probability mass function as objective function in QMLE is also motivated by Gourieroux, Monfort, and Trognon (1984a, b), who demonstrate that the negative binomial distribution with fixed value of δ is in the linear exponential family (LEF) of distributions. Gourieroux, Monfort, and Trognon (1984a, b) show that QMLE is a consistent estimator for LEF, provided that the conditional mean of the dependent variable is correctly specified. For the pooled negative binomial QMLE we assume that

(QMLE1) The conditional mean of patent count is correctly specified, i.e., $E(n_{it}|X_{i1}, \ldots, X_{it}) = \lambda(X_{it}, \theta) = \lambda_{it}$. This implies the weak exogeneity (Cameron and Trivedi (2005)) of all explanatory variables.

We implement the QMLE procedure (Gourieroux, Monfort, and Trognon (1984a, b)) following the two-step approach suggested by Wooldridge (1997a, Sections 4.2 and 9.2). In the first step, the δ parameter of the negative binomial distribution is estimated. In the second step, $\hat{\delta}$ is included into LL of the negative binomial distribution and QMLE is performed to estimate θ. The separate estimation of δ and θ is motivated by Gourieroux, Monfort, and Trognon (1984a, b). The details of the two-step QMLE negative binomial procedure are as follows. In the first step, the quasi-log-likelihood objective function for the pooled Poisson estimation is

$$(3.28)\quad LL(X_{i1}, \ldots, X_{iT}, \theta) = \sum_{i=1}^{N}\sum_{t=1}^{T} n_{it} \ln \lambda(X_{it}, \theta) - \lambda(X_{it}, \theta) = \sum_{i=1}^{N}\sum_{t=1}^{T} n_{it} \ln(\lambda_{it}) - \lambda_{it}.$$

The pooled Poisson QMLE, denoted by $\hat{\theta}$, maximizes LL. Gourieroux, Monfort, and Trognon (1984a, b) show that the Poisson distribution is LEF, hence the Poisson QMLE is a consistent

estimator under correct specification of the conditional mean of patent count. We define the Poisson residuals by $\hat{u}_{it} = n_{it} - \lambda(X_{it}, \hat{\theta})$, and also define the weighted (or Pearson) residuals by $\tilde{u}_{it} = \hat{u}_{it}/\sqrt{\lambda(X_{it}, \hat{\theta})}$. Given these residuals, we estimate δ by pooled ordinary least squares (OLS) for the following linear regression model (Wooldridge, 1997a, Section 9.2):

$$(3.29)\quad \tilde{u}_{it}^2 - 1 = c + \delta\lambda(X_{it}, \hat{\theta}) + \epsilon_{it}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The pooled OLS provides $\hat{\delta}$, which we substitute into LL of the second step. In the second step, the quasi-log-likelihood objective function for pooled negative binomial estimation is

$$(3.30)\quad LL(X_{i1}, \ldots, X_{iT}, \hat{\delta}, \theta) = \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\delta}^{-1} \ln\left[\frac{\hat{\delta}^{-1}}{\hat{\delta}^{-1} + \lambda(X_{it}, \theta)}\right] + n_{it} \ln\left[\frac{\lambda(X_{it}, \theta)}{\hat{\delta}^{-1} + \lambda(X_{it}, \theta)}\right].$$

The pooled negative binomial QMLE, denoted by $\hat{\theta}$, maximizes LL. Gourieroux, Monfort, and Trognon (1984a, b) show that the negative binomial distribution is LEF. Therefore, the pooled negative binomial QMLE is consistent under correct specification of the conditional mean of patent count.

The asymptotic variance of $\hat{\theta}$ is estimated by the following robust estimator. The asymptotic variance of $\hat{\theta}$ is given by $A^{-1}BA^{-1}/N$. This is due to the following result demonstrated by Wooldridge (1997a): $\sqrt{N}(\hat{\theta} - \theta) \to_d N(0, A^{-1}BA^{-1})$, where

$$(3.31)\quad A = \sum_{t=1}^{T} E\left[\frac{\nabla_\theta \lambda(X_{it}, \theta)' \nabla_\theta \lambda(X_{it}, \theta)}{\lambda(X_{it}, \theta)}\right]$$

$$(3.32)\quad B = E[s_i(\theta) s_i(\theta)'].$$

In the last expression, $s_i(\theta) = \sum_{t=1}^{T} s_{it}(\theta)$ and the score $s_{it}(\theta)$ is

$$(3.33)\quad s_{it}(\theta) = \frac{\nabla_\theta \lambda(X_{it}, \theta)'[n_{it} - \lambda(X_{it}, \theta)]}{\lambda(X_{it}, \theta) + \hat{\delta}\lambda(X_{it}, \theta)^2}.$$

For a panel with randomly sampled cross-section, consistent estimators of A and B are

$$(3.34)\quad \hat{A} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\nabla_\theta\lambda(X_{it},\hat{\theta})'\nabla_\theta\lambda(X_{it},\hat{\theta})}{\lambda(X_{it},\hat{\theta})+\hat{\delta}\lambda(X_{it},\hat{\theta})^2}$$

$$(3.35)\quad \hat{B} = N^{-1}\sum_{i=1}^{N} s_i(\hat{\theta})s_i(\hat{\theta})'.$$

All count panel data models of this paper can be consistently estimated by the pooled negative binomial QMLE method, given that the conditional mean of patent count is correctly specified.

3.7. *Parameter Estimation by GMM*

Chamberlain (1992) and Wooldridge (1997b) use the GMM method for count panel data models with unobserved effects. These authors use GMM for a transformation of patent counts for which the GMM moment conditions hold. Both authors suggest the GMM method for those cases when strict exogeneity of explanatory variables fails. Examples of these cases are the dynamic count panel data models with feedback effects. For GMM we assume that

(GMM1) The conditional mean of n_{it} is correctly specified, $E(n_{it}|X_{i1},\ldots,X_{it},\alpha_i)=\lambda(X_{it},\theta)$. This implies the weak exogeneity of all explanatory variables, conditional on α_i. Chamberlain (1992) refers to this as sequential moment restrictions (Wooldridge (1997a), Section 10.2; Wooldridge (1997b)).

(GMM2) The conditional mean function is the exponential function, $\lambda_{it}=\exp(X_{it}\beta+\ln\alpha_i)$.

The (GMM1) assumption is weaker than (ML1) since strict exogeneity is not required. The information set in (GMM1) and the model formulation in (GMM2) includes the unobserved effect parameter α_i. Nevertheless, we do not restrict the distribution of α_i conditional on the explanatory variables. (GMM2) coincides with (ML2), nevertheless (GMM2) may be relaxed to consider different functional forms of patent conditional expectation (e.g., additive functional forms). For example, Blundell, Griffith, and Windmeijer (2002) introduce the additive LFM

and demonstrate the corresponding moment conditions. Furthermore, the parameters of time-constant explanatory variables are not identified by GMM for the exponential count data model (Wooldridge (1997b)). This property is similar to FE MLE. Therefore, GMM can be seen as an alternative of FE MLE with weaker maintained assumptions. Under (GMM2), GMM is applied to the following transformation of the dependent variable, for each firm $i = 1, \ldots, N$ and period $t = 1, \ldots, T-1$:

$$(3.36)\quad r_{it}(\theta) = r_{it} = n_{it} - n_{it+1}\frac{\exp(X_{it}\beta)}{\exp(X_{it+1}\beta)},$$

where $r_{it}(\theta)$ indicates that the transformed variable depends on the vector of parameters θ. Wooldridge (1997b) demonstrates that under (GMM1) and (GMM2), $E(r_{it}|X_{i1}, \ldots, X_{it}, \alpha_i) = 0$ which is the basis for the GMM estimator. For firm i, we introduce the $(T-1)\times 1$ vector notation $r_i = (r_{i1}, \ldots, r_{iT-1})'$ for the transformed dependent variables and we also introduce the following notation for the matrix of instrumental variables:

$$(3.37)\quad Z_i = \begin{bmatrix} z_{i1} & 0 & 0 & \cdots & 0 \\ 0 & z_{i2} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ 0 & \cdots & & 0 & z_{iT-1} \end{bmatrix}.$$

The general element z_{it} of this matrix is a vector with dimensions $1 \times L_t$. We choose the instrumental variables in Z_i as follows. First, $z_{i1} = (1, rd_{i1})$ thus $L_1 = 2$. Then, a general element of Z_i is given by $z_{it} = (1, rd_{i1}, \ldots, rd_{it}, n_{i1}, \ldots, n_{it-1})$. This implies that the number of elements of L_t is increasing with t and $L_{t+1} = L_t + 2$. The dimensions of Z_i are $(T-1) \times L$, where $L = L_1 + \ldots + L_{T-1}$ is the number of all instrumental variables. For our dataset and variables the dimensions of Z_i is 21×462. We use the same matrix of instrumental variables for

all models. The GMM estimator is given by

$$(3.38)\quad \hat{\theta} = \arg\min_{\theta} \left[\sum_{i=1}^{N} Z_i' r_i(\theta)\right]' W \left[\sum_{i=1}^{N} Z_i' r_i(\theta)\right],$$

where W denotes the $L \times L$ general positive definite weight matrix. There is no closed form solution to this problem, since the expression to be minimized is non-linear in θ. Therefore, we solve it numerically. For the weight matrix we use

$$(3.39)\quad W = \hat{\Omega}^{-1} = \left[N^{-1} \sum_{i=1}^{N} Z_i' r_i r_i' Z_i\right]^{-1}.$$

The asymptotic distribution and the robust covariance matrix of parameter estimates is obtained by the following result (Wooldridge (1997b)):

$$(3.40)\quad \sqrt{N}(\hat{\theta} - \theta) \sim_a N\left[0, \left(R'\Omega^{-1}R\right)^{-1}\right],$$

where Ω^{-1} is estimated according to equation (3.39) and R is an $L \times K$ matrix (K is the number of parameters). Moreover, R is estimated as

$$(3.41)\quad \hat{R} = N^{-1} \sum_{i=1}^{N} Z_i' \frac{\partial r_{it}(\hat{\theta})}{\partial \theta},$$

where $\partial r_{it}(\hat{\theta})/\partial\theta$ for the parameter $\tilde{\beta} \in \beta$ that corresponds to $x_{it} \in X_{it}$ is given by

$$(3.42)\quad \frac{\partial r_{it}(\theta)}{\partial \tilde{\beta}} = -n_{it+1}(x_{it} - x_{it+1}) \frac{\exp(X_{it}\beta)}{\exp(X_{it+1}\beta)}.$$

We combine equations (3.36) and (3.42), to obtain

$$(3.43)\quad \frac{\partial r_{it}(\theta)}{\partial \tilde{\beta}} = (r_{it} - n_{it})(x_{it} - x_{it+1}).$$

For time-constant explanatory variables equation (3.43) is zero for all t, hence the estimate of

$R'\Omega^{-1}R$ is a singular matrix. Therefore, GMM standard errors cannot be computed for models with time-constant explanatory variables. The asymptotic covariance matrix of $\hat{\theta}$ is estimated by $(\hat{R}'\hat{\Omega}^{-1}\hat{R})^{-1}/N$, and it is evaluated at the GMM parameter estimates $\hat{\theta}$.

The joint null hypothesis of adequate functional form of λ_{it} and exogeneity of instrumental variables can be tested by the GMM overidentification test statistic, evaluated at the GMM parameter estimates (Hansen (1982); Wooldridge (1997b)):

$$(3.44)\quad N^{-1}\left[\sum_{i=1}^{N} Z_i' r_i(\hat{\theta})\right]' \hat{\Omega}^{-1} \left[\sum_{i=1}^{N} Z_i' r_i(\hat{\theta})\right] \sim_a \chi^2(L-K).$$

Although GMM is very general with few assumptions maintained, we do not use this method to estimate parameters of the count panel data models due to the following reasons. First, GMM assumes that the unobserved effect α_i appears in λ_{it}. If lags of the conditional score are included in $X_{it}\beta$, as in both DCS models, then r_{it} cannot be computed due to the latent α_i term. This makes the GMM distance minimization problem unfeasible for the DCS count panel data models. Similar to MLE, this issue motivates the application of pooled count panel data models. Second, the GMM numerical estimation procedure was very slow for our dataset. In order to increase the speed of GMM we used a reduced number of instrumental variables, as suggested by Wooldridge (1997b, p. 675). In this way the dimensions of the matrix of instrumental variables Z_i are reduced to 21×82, and hence the speed of the GMM code increased significantly. However, these instruments failed the GMM overidentification test of equation (3.44), and we also had numerical problems related to the singularity of $\hat{\Omega}$ for the GMM procedure. As a consequence, the GMM minimization problem did not converge effectively.

4. MODEL DIAGNOSTICS AND EMPIRICAL RESULTS

In this section we present the diagnostic tests and empirical results for the pooled negative binomial QMLE method. Table I shows the parameter estimates of the following multiplicative patent count panel data models: FDL, EFM, MDCS-QMA(5), MDCS-QAR(1), MDCS-QARMA(1,1) and MDCS-QARMA(1,5). Table II shows the parameter estimates of the follow-

ing additive patent count panel data models: LFM, ADCS-QMA(5), ADCS-QAR(1), ADCS-QARMA(1,1) and ADCS-QARMA(1,5). In both tables we report robust standard errors of the parameters, obtained by the robust sandwich covariance matrix estimator (Davidson and MacKinnon (2003)). The last row of Tables I and II presents the pooled OLS estimate of δ for each model for the first step of the pooled negative binomial QMLE procedure. Other rows of Tables I and II show the second step of the pooled negative binomial QMLE. Tables I and II present the following interesting results. First, the parameter estimates of $\gamma_1, \ldots, \gamma_5$ and their significance are similar for all count data models. Second, for LFM the dynamic coefficient is not far from one, $\hat{\phi}_1 = 0.92$, hence the patent count process almost has a unit root. Third, for LFM besides the contemporaneous R&D effect all R&D effects are negative. This result is strange and hence questions the consistency of parameter estimates for LFM. Fourth, regarding $\hat{\delta}$ of LFM we can see in Table II that this parameter is very low, with respect to all other count data models. Fifth, for both ADCS formulations the estimates of δ and all R&D effects are similar to those of MDCS count panel data models.

If the conditional mean of patent count is correctly specified, then contemporaneous R&D expenditure will be an exogenous variable. Nevertheless, R&D expenses are possibly simultaneous with patent application count for all patent count panel data models. We test the exogeneity of R&D expenses according to Wooldridge (1997a, Section 6.1) and Wooldridge (2002, Section 19.5.1). For all models we consider that contemporaneous log R&D expenses rd_{it} and the interaction term $(t \times rd_{it})$ are potentially endogenous, while other variables in X_{it} are exogenous. Let Z_{it} denote the exogenous variables in X_{it}. For different models, Z_{it} is

(4.1) FDL: $Z_{it} = (1, D_i, b_i, rd_{it-1}, \ldots, rd_{it-k})$

(4.2) EFM and LFM: $Z_{it} = (1, n_{it-1}, n_{i1}, D_i, b_i, rd_{it-1}, \ldots, rd_{it-k})$

(4.3) MDCS and ADCS: $Z_{it} = (1, u_{it-1}, \ldots, u_{it-q}, n_{i1}, D_i, b_i, rd_{it-1}, \ldots, rd_{it-k})$.

We test the exogeneity of R&D expenses by the two-step approach of Wooldridge (1997a, 2002),

applied to count panel data model. In the first step we obtain the cross-sectional OLS estimates from the following regressions, for each period $t = 1, \ldots, T$:

(4.4) $(t \times rd_{it}) = \psi_1 + Z_{it}\Pi_1 + v_{1it}$

(4.5) $rd_{it} = \psi_2 + Z_{it}\Pi_2 + v_{2it}$

with $i = 1, \ldots, N$, v_{1it} and v_{2it} denote error terms. Denote the OLS residuals by $\hat{v}_{1it}$ and $\hat{v}_{2it}$. In the second step we include these residuals into the extended panel data model:

(4.6) FDL and EFM: $\lambda_{it} = \exp(X_{it}\theta + \hat{v}_{1it}\rho_1 + \hat{v}_{2it}\rho_2)$

(4.7) MDCS: $\lambda_{it} = \exp(\Psi_{it} + Y_{it}\tilde{\theta} + \hat{v}_{1it}\rho_1 + \hat{v}_{2it}\rho_2)$

(4.8) LFM: $\lambda_{it} = \phi_1 n_{it-1} + \exp(Y_{it}\tilde{\theta} + \hat{v}_{1it}\rho_1 + \hat{v}_{2it}\rho_2)$

(4.9) ADCS: $\lambda_{it} = \Psi_{it} + \exp(Y_{it}\tilde{\theta} + \hat{v}_{1it}\rho_1 + \hat{v}_{2it}\rho_2)$.

With respect to the coefficients ρ_1 and ρ_2, the null hypothesis that both rd_{it} and $(t \times rd_{it})$ are exogenous is equivalent with $H_0 : (\rho_1 = 0 \text{ and } \rho_2 = 0)$. We use the robust QMLE to test if ρ_1 or ρ_2 are significantly different from zero. If they are non-significant, then we will conclude that there is no evidence against the hypothesis that R&D expenses are exogenous. Panels A and B of Table III present the pooled negative binomial QMLE of ρ_1 or ρ_2, with robust standard errors. Table III demonstrates that the null hypothesis according to which R&D is exogenous cannot be rejected at the 1% level of significance for FDL, EFM, MDCS-QARMA(1,5) and ADCS-QARMA(1,5). For other count panel data models R&D is an endogenous explanatory variable according to the test, thus the conditional mean of patent count is not specified correctly and the pooled negative binomial QMLE is not a consistent estimator for these models. The estimation and test results show that for MDCS and ADCS, QAR and several QMA lags are needed to make R&D expenses exogenous in the conditional mean equation.

We compare the statistical performance of different patent count panel data models by R-squared-type model performance metrics. Cameron and Windmeijer (1996) suggest two R-squared metrics for count data models with negative binomial specification of the second kind. Cameron and Windmeijer (1996) present the R-squared formulas for the cross-sectional data case. We implement these formulas for the panel data setup by computing each R-squared for each time period. The first one is the Pearson R-squared,

$$(4.10)\quad R^2_{\mathrm{P,NB2},t} = 1 - \frac{\sum_{i=1}^{N}(n_{it}-\lambda_{it})^2/(\lambda_{it}+\hat{\delta}\lambda_{it}^2)}{\sum_{i=1}^{N}(n_{it}-\overline{n}_t)^2/(\overline{n}_t+\hat{\delta}\overline{n}_t^2)}$$

and the second R-squared is based on deviance residuals for ML estimation,

$$(4.11)\quad R^2_{\mathrm{DEV,NB2}(ML),t} = 1 - \frac{\sum_{i=1}^{N}\left[n_{it}\ln\left(\frac{n_{it}}{\lambda_{it}}\right) - (n_{it}+\hat{\delta}^{-1})\ln\left(\frac{n_{it}+\hat{\delta}^{-1}}{\lambda_{it}+\hat{\delta}^{-1}}\right)\right]}{\sum_{i=1}^{N}\left[n_{it}\ln\left(\frac{n_{it}}{\lambda_{it}}\right) - (n_{it}+\hat{\delta}^{-1})\ln\left(\frac{n_{it}+\hat{\delta}^{-1}}{\overline{n}_t+\hat{\delta}^{-1}}\right)\right]}.$$

Both R-squared values are defined for each period $t = 1, \ldots, T$. In Panels C and D of Table III we present the simple average over time of Pearson R-squared and deviance residual R-squared series. We show by bold numbers those well-specified models for which R&D is an exogenous variable. The results show that MDCS-QARMA(1,5) has the highest mean R-squared values, with respect to those models where R&D is exogenous. Due to the fact that simple average may be an inconsistent estimator and also misleading, we also present the evolution of Pearson R-squared and deviance R-squared in Figures 1 and 2, respectively, for period 1979 to 2000. In these figures, we present only those models for which R&D is exogenous. Figures 1 and 2 show that the superiority of MDCS-QARMA(1,5) is persistent over time, according to both R-squared metrics.

5. CONCLUSIONS

In this paper we have introduced a new class of DCS models that allows us to estimate DCS time-series models for firm-level panels of patent count data. We have estimated several patent count panel data models for the extended panel data of patent applications used by

Blazsek and Escribano (2010, 2015). Different patent count data models have been estimated for a large panel of 4,476 US firms for period 1979 to 2000. We have considered three alternative estimation methods, MLE, QMLE and GMM, for count panel data models, and we conclude that the negative binomial QMLE is the most appropriate method. Exogeneity tests have indicated that R&D is exogenous for FDL, EFM, MDCS-QARMA(1,5) and ADCS-QARMA(1,5). For all other count panel data models R&D expenditure seems to be an endogenous variable, mainly due to omitted variables and the fact that the models are not dynamically complete. Hence, the conditional mean of patent count is misspecified. We have also used different R-squared metrics in order to compare the statistical performance of count panel data models, which suggest that the MDCS-QARMA(1,5) model is superior to other count panel data models considered.

REFERENCES

Blazsek, S., and A. Escribano (2010): "Knowledge Spillovers in U.S. Patents: A Dynamic Patent Intensity Model with Secret Common Innovation Factors," *Journal of Econometrics*, 159 (1), 14–32.

Blazsek, S., and A. Escribano (2015): "Patent Propensity, R&D and Market Competition: Dynamic Spillovers of Innovation Leaders and Followers," *Journal of Econometrics*, http://dx.doi.org/10.1016/j.jeconom.2015.10.005

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81 (4), 1347–1393.

Blundell, R., R. Griffith, and F. Windmeijer (2002): "Individual Effects and Dynamics in Count Data Models," *Journal of Econometrics*, 108 (1), 113–131.

Cameron, A. C., and P. K. Trivedi (1986): "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics*, 1 (1), 29–53.

Cameron, A. C., and P. K. Trivedi (2005): *Microeconometrics Methods and Applications*. Cambridge, UK: Cambridge University Press.

Cameron, A. C., and F. A. G. Windmeijer (1996): "R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization," *Journal of Business & Economic Statistics*, 14 (2), 209–220.

Chamberlain, G. (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60 (3), 567–596.

Davidson, R., and J. G. MacKinnon (2003): *Econometric theory and methods*. New York: Oxford University Press.

Davis, R., W. Dunsmuir, and S. Streett (2003): "Observation-Driven Models for Poisson Counts," *Biometrika*, 90 (4), 777–790.

Davis, R., W. Dunsmuir, and S. Streett (2005): "Maximum Likelihood Estimation for an Observation Driven Model for Poisson Counts," *Methodology & Computing in Applied Probability*, 7 (2), 149–159.

Gourieroux, C., A. Monfort, and A. Trognon (1984a): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52 (3), 681–700.

Gourieroux, C., A. Monfort, and A. Trognon (1984b): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52 (3), 701–720.

Griliches, Z. (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28 (4), 1661–1707.

Hall, B., A. B. Jaffe, and M. Trajtenberg (2001): "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper No. 8498.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 (4), 1029–1054.

Harvey, A. C. (2013): *Dynamic Models for Volatility and Heavy Tails*. Cambridge, UK: Cambridge University Press.

Hausman, J., B. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52 (4), 909–938.

Jaffe, A. B. (1986): "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, 76 (5), 984–1001.

Lanjouw, J. O., A. Pakes, and J. Putnam (1998): "How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data," *The Journal of Industrial Economics*, 46 (4), 405–432.

Lanjouw, J. O., and M. Schankerman (1999): "The Quality of Ideas: Measuring Innovation with Multiple Indicators," NBER Working Paper No. 7345.

Pakes, A. (1985): "On Patents, R&D, and the Stock Market Rate of Return," *Journal of Political Economy*, 93 (2), 390–409.

Trajtenberg, M. (1990): "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, 21 (1), 172–187.

Wooldridge, J. M. (1997a): "Quasi-Likelihood Methods for Count Data," in *Handbook of Applied Econometrics*, Volume 2, ed. by M. H. Pesaran and P. Schmidt. Oxford: Blackwell, 352–406.

Wooldridge, J. M. (1997b): "Multiplicative Panel Data Models without the Strict Exogeneity Assumption," *Econometric Theory*, 13 (5), 667–678.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

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.

TABLE I

MULTIPLICATIVE COUNT PANEL DATA MODELS: POOLED NEGATIVE BINOMIAL QMLE PARAMETER ESTIMATES

Parameter	FDL	EFM	MDCS-QMA(1)	MDCS-QAR(1)	MDCS-QARMA(1,1)	MDCS-QARMA(1,5)
μ_0	$-1.8050^{***}(0.1838)$	$-1.7523^{***}(0.1542)$	$-1.1213^{***}(0.0392)$	$-1.2426^{***}(0.0309)$	$-1.2449^{***}(0.0308)$	$-1.2505^{***}(0.0304)$
$\gamma_1 t$	$0.0757^{***}(0.0062)$	$0.0762^{***}(0.0054)$	$0.0397^{***}(0.0015)$	$0.0594^{***}(0.0016)$	$0.0598^{***}(0.0016)$	$0.0607^{***}(0.0016)$
$\gamma_2(t \times rd_{it})$	$-0.0245^{***}(0.0030)$	$-0.0194^{***}(0.0025)$	$-0.0080^{***}(0.0011)$	$-0.0079^{***}(0.0011)$	$-0.0079^{***}(0.0011)$	$-0.0079^{***}(0.0011)$
$\gamma_3 D_i$	$0.2133^{**}(0.0910)$	$0.2476^{***}(0.0731)$	$0.1835^{***}(0.0314)$	$0.1583^{***}(0.0279)$	$0.1577^{***}(0.0278)$	$0.1599^{***}(0.0275)$
$\gamma_4 z_i$	$0.1168^{***}(0.0344)$	$0.0996^{***}(0.0273)$	$0.0653^{***}(0.0075)$	$0.0572^{***}(0.0054)$	$0.0568^{***}(0.0054)$	$0.0559^{***}(0.0052)$
$\gamma_5 n_{i1}$	NA	$0.0066^{***}(0.0014)$	$0.0157^{***}(0.0007)$	$0.0161^{***}(0.0007)$	$0.0161^{***}(0.0007)$	$0.0161^{***}(0.0007)$
$\beta_0 rd_{it}$	$0.8631^{***}(0.0602)$	$0.7394^{***}(0.0525)$	$0.3804^{***}(0.0296)$	$0.3775^{***}(0.0309)$	$0.3776^{***}(0.0309)$	$0.3784^{***}(0.0312)$
$\beta_1 rd_{it-1}$	$0.0502^{**}(0.0218)$	$0.0136(0.0172)$	$0.0294(0.0228)$	$0.0247(0.0232)$	$0.0251(0.0234)$	$0.0241(0.0234)$
$\beta_2 rd_{it-2}$	$0.0492^{**}(0.0227)$	$0.0383^{***}(0.0127)$	$0.0372^{***}(0.0097)$	$0.0353^{***}(0.0094)$	$0.0356^{***}(0.0095)$	$0.0369^{***}(0.0094)$
$\beta_3 rd_{it-3}$	$0.0965^{***}(0.0353)$	$0.0650^{***}(0.0166)$	$0.0429^{***}(0.0104)$	$0.0288^{***}(0.0100)$	$0.0284^{***}(0.0101)$	$0.0291^{***}(0.0101)$
$\beta_4 rd_{it-4}$	$0.0165(0.0152)$	$-0.0061(0.0113)$	$0.0368^{***}(0.0114)$	$0.0196^{*}(0.0106)$	$0.0189^{*}(0.0106)$	$0.0192^{*}(0.0107)$
$\beta_5 rd_{it-5}$	$0.0794^{***}(0.0265)$	$-0.0245(0.0192)$	$0.0587^{***}(0.0149)$	$0.0379^{***}(0.0130)$	$0.0375^{***}(0.0129)$	$0.0356^{***}(0.0131)$
ϕ_1 AR(1)	NA	$0.0113^{***}(0.0003)$	NA	$0.8682^{***}(0.0119)$	$0.8721^{***}(0.0134)$	$0.8992^{***}(0.0238)$
$\theta_0 u_{it}$	NA	NA	$0.2390^{***}(0.0026)$	$0.1968^{***}(0.0025)$	$0.2000^{***}(0.0025)$	$0.1999^{***}(0.0025)$
$\theta_1 u_{it-1}$	NA	NA	$0.2008^{***}(0.0027)$	NA	$-0.0066^{***}(0.0020)$	$-0.0079^{*}(0.0048)$
$\theta_2 u_{it-2}$	NA	NA	$0.1653^{***}(0.0029)$	NA	NA	$-0.0069^{*}(0.0042)$
$\theta_3 u_{it-3}$	NA	NA	$0.1342^{***}(0.0028)$	NA	NA	$-0.0064^{*}(0.0036)$
$\theta_4 u_{it-4}$	NA	NA	$0.1037^{***}(0.0028)$	NA	NA	$-0.0049(0.0031)$
$\theta_5 u_{it-5}$	NA	NA	$0.0320^{**}(0.0134)$	NA	NA	$-0.0089^{***}(0.0027)$
δ	$0.8008^{***}(0.1401)$	$0.3111^{***}(0.0822)$	$0.3120^{***}(0.0549)$	$0.2593^{***}(0.0386)$	$0.2595^{***}(0.0386)$	$0.2587^{***}(0.0388)$

Notes: Robust standard errors are reported in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

TABLE II

ADDITIVE COUNT PANEL DATA MODELS: POOLED NEGATIVE BINOMIAL QMLE PARAMETER ESTIMATES

Parameter	LFM	ADCS-QMA(1)	ADCS-QAR(1)	ADCS-QARMA(1,1)	ADCS-QARMA(1,5)
μ_0	$-2.5382^{***}(0.0421)$	$-2.2299^{***}(0.0851)$	$-2.2377^{***}(0.0848)$	$-2.2458^{***}(0.0905)$	$-2.2519^{***}(0.0907)$
$\gamma_1 t$	$0.0394^{***}(0.0018)$	$0.0323^{***}(0.0030)$	$0.0301^{***}(0.0030)$	$0.0302^{***}(0.0030)$	$0.0296^{***}(0.0031)$
$\gamma_2(t \times rd_{it})$	$-0.0160^{***}(0.0006)$	$-0.0076^{***}(0.0021)$	$-0.0072^{***}(0.0021)$	$-0.0071^{***}(0.0021)$	$-0.0069^{***}(0.0021)$
$\gamma_3 D_i$	$0.3545^{***}(0.0146)$	$0.3209^{***}(0.0493)$	$0.3232^{***}(0.0492)$	$0.3248^{***}(0.0496)$	$0.3301^{***}(0.0497)$
$\gamma_4 z_i$	$0.0940^{***}(0.0037)$	$0.1178^{***}(0.0153)$	$0.1195^{***}(0.0150)$	$0.1203^{***}(0.0152)$	$0.1224^{***}(0.0151)$
$\gamma_5 n_{i1}$	$0.0059^{***}(0.0003)$	$0.0092^{***}(0.0009)$	$0.0091^{***}(0.0009)$	$0.0090^{***}(0.0009)$	$0.0089^{***}(0.0009)$
$\beta_0 rd_{it}$	$1.0647^{***}(0.0148)$	$0.6468^{***}(0.0563)$	$0.6461^{***}(0.0569)$	$0.6442^{***}(0.0568)$	$0.6430^{***}(0.0571)$
$\beta_1 rd_{it-1}$	$-0.3582^{***}(0.0104)$	$0.1192^{***}(0.0366)$	$0.1164^{***}(0.0364)$	$0.1200^{***}(0.0369)$	$0.1178^{***}(0.0373)$
$\beta_2 rd_{it-2}$	$-0.0834^{***}(0.0079)$	$0.0590^{***}(0.0151)$	$0.0589^{***}(0.0156)$	$0.0578^{***}(0.0151)$	$0.0573^{***}(0.0151)$
$\beta_3 rd_{it-3}$	$-0.0206^{**}(0.0089)$	$0.0369^{**}(0.0177)$	$0.0436^{**}(0.0174)$	$0.0430^{**}(0.0182)$	$0.0444^{**}(0.0176)$
$\beta_4 rd_{it-4}$	$-0.0286^{***}(0.0109)$	0.0101(0.0099)	0.0053(0.0097)	0.0045(0.0097)	0.0050(0.0098)
$\beta_5 rd_{it-5}$	$-0.0373^{***}(0.0075)$	$0.0749^{***}(0.0209)$	$0.0783^{***}(0.0208)$	$0.0790^{***}(0.0207)$	$0.0782^{***}(0.0207)$
ϕ_1 AR(1)	$0.9215^{***}(0.0254)$	NA	$0.5746^{***}(0.0221)$	$0.6159^{***}(0.0404)$	$0.8328^{***}(0.0478)$
$\theta_0 u_{it}$	NA	$0.3690^{***}(0.0065)$	$0.3292^{***}(0.0065)$	$0.3567^{***}(0.0083)$	$0.3586^{***}(0.0107)$
$\theta_1 u_{it-1}$	NA	$0.1877^{***}(0.0032)$	NA	$-0.0569^{***}(0.0100)$	$-0.1180^{***}(0.0094)$
$\theta_2 u_{it-2}$	NA	$0.0926^{***}(0.0024)$	NA	NA	$-0.0606^{***}(0.0043)$
$\theta_3 u_{it-3}$	NA	$0.0536^{***}(0.0021)$	NA	NA	$-0.0238^{***}(0.0025)$
$\theta_4 u_{it-4}$	NA	$0.0327^{***}(0.0020)$	NA	NA	$-0.0091^{***}(0.0021)$
$\theta_5 u_{it-5}$	NA	$0.0130^{***}(0.0017)$	NA	NA	$-0.0128^{***}(0.0020)$
δ	$0.0476^{***}(0.0069)$	$0.5644^{***}(0.1277)$	$0.5641^{***}(0.1282)$	$0.5637^{***}(0.1282)$	$0.5631^{***}(0.1281)$

Notes: Robust standard errors are reported in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively.

TABLE III

Exogeneity Tests and Model Performance Metrics

Panel A. Test of exogeneity of R&D for multiplicative models						
Parameter	FDL	EFM	MDCS-QMA(1)	MDCS-QAR(1)	MDCS-QARMA(1,1)	MDCS-QARMA(1,5)
ρ_1	0.0012(0.0098)	−0.0053(0.0069)	−0.0016(0.0047)	0.0018(0.0047)	0.0017(0.0047)	0.0018(0.0585)
ρ_2	−0.2337(0.1774)	−0.1940(0.1240)	−0.2284***(0.0882)	−0.2590***(0.0915)	−0.2582***(0.0910)	−0.2582(1.2093)
Test result	R&D exogenous	R&D exogenous	R&D endogenous	R&D endogenous	R&D endogenous	R&D exogenous
Panel B. Test of exogeneity of R&D for additive models						
Parameter		LFM	ADCS-QMA(1)	ADCS-QAR(1)	ADCS-QARMA(1,1)	ADCS-QARMA(1,5)
ρ_1		0.0054**(0.0027)	−0.0013(0.0104)	−0.0011(0.0105)	−0.0012(0.0104)	−0.0013(0.0842)
ρ_2		−0.7617***(0.0492)	−0.5817***(0.1537)	−0.5968***(0.1542)	−0.5956***(0.1546)	−0.5932(1.7437)
Test result		R&D endogenous	R&D endogenous	R&D endogenous	R&D endogenous	R&D exogenous
Panel C. Performance metrics for multiplicative models						
Mean R^2	FDL	EFM	MDCS-QMA(1)	MDCS-QAR(1)	MDCS-QARMA(1,1)	MDCS-QARMA(1,5)
$\overline{R}^2_{\text{P,NB2}}$	**84.90**%	**94.53**%	98.67%	98.70%	98.70%	**98.69**%
$\overline{R}^2_{\text{DEV,NB2(ML)}}$	**61.21**%	**71.10**%	81.07%	82.57%	82.57%	**82.59**%
Panel D. Performance metrics for additive models						
Mean R^2		LFM	ADCS-QMA(1)	ADCS-QAR(1)	ADCS-QARMA(1,1)	ADCS-QARMA(1,5)
$\overline{R}^2_{\text{P,NB2}}$		97.73%	96.57%	96.54%	96.53%	**96.52**%
$\overline{R}^2_{\text{DEV,NB2(ML)}}$		89.92%	74.63%	74.71%	74.72%	**74.74**%

Notes: We study whether R&D is exogenous by using the Wooldridge (1997a, 2002) test. Significant ρ_1 or ρ_2 indicate endogeneity of current R&D expenditure. Robust standard errors are reported in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively. Pearson R-squared and deviance residual R-squared values are computed according to Cameron and Windmeijer (1996) and applied to panel data. Bold R-squared values indicate that R&D is exogenous variable for the corresponding model.

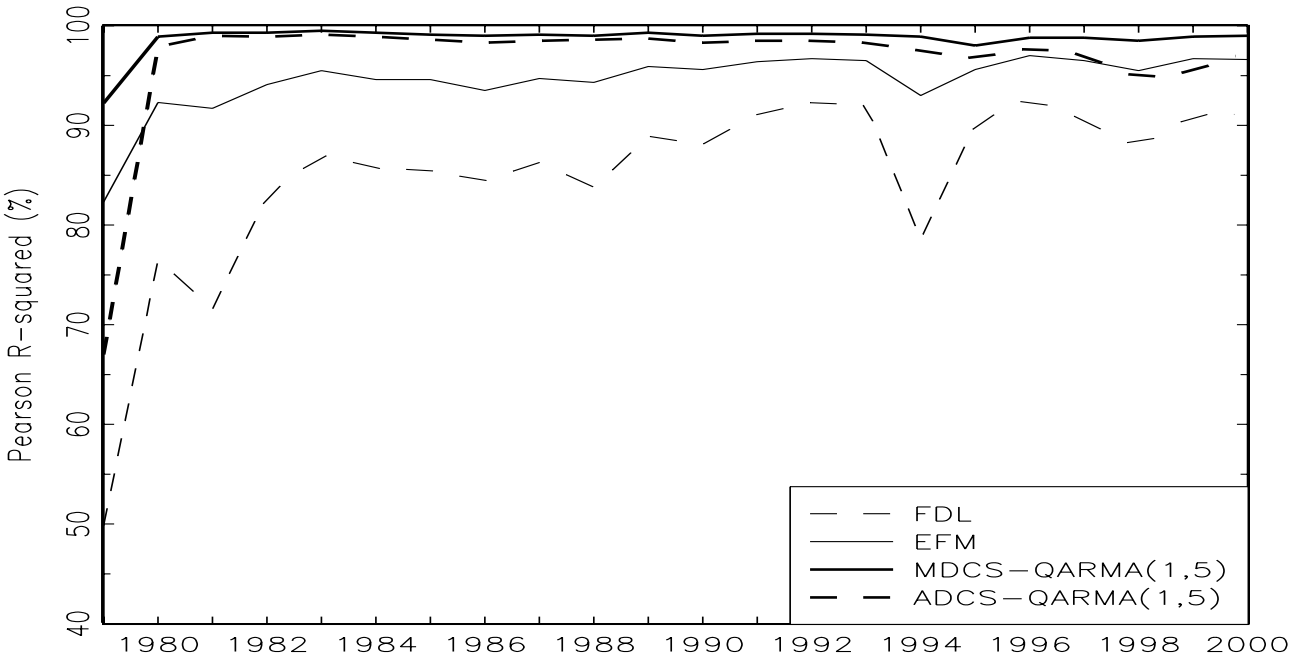

FIGURE 1.—Pearson R-squared for period 1979 to 2000.

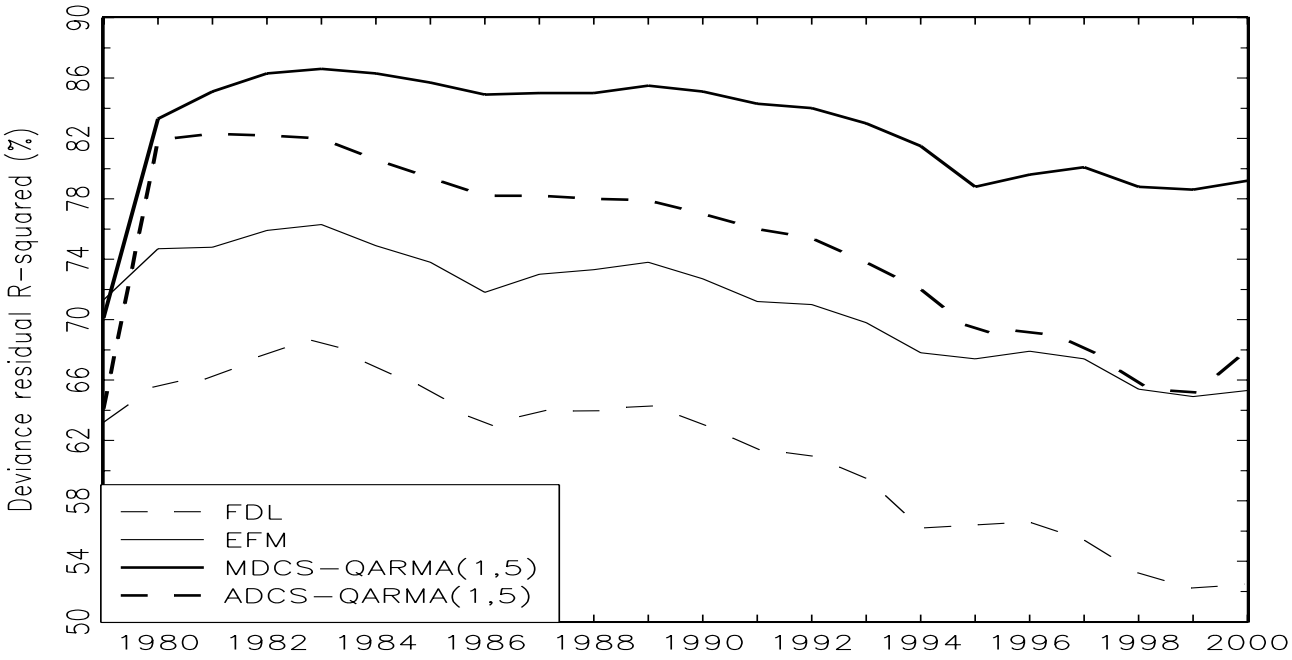

FIGURE 2.—Deviance residual R-squared for period 1979 to 2000.