INDIVIDUAL EFFECTS AND DYNAMICS IN COUNT DATA MODELS

Richard Blundell
Rachel Griffith
Frank Windmeijer

Individual Effects and Dynamics in Count Data Models[*]

Richard Blundell[†]   Rachel Griffith[‡]   Frank Windmeijer[§]

June 19, 2000

Abstract

In this paper we examine the panel data estimation of dynamic models for count data that include correlated fixed effects and predetermined variables. Use of a linear feedback model is proposed. The standard Poisson conditional maximum likelihood estimator for non-dynamic models, which is shown to be the same as the Poisson maximum likelihood estimator in a model with individual specific constants, is inconsistent when regressors are predetermined. A quasi-differenced GMM estimator is consistent for the parameters in the dynamic model, but when series are highly persistent, there is a problem of weak instrument bias. An estimator is proposed that utilises pre-sample information of the dependent count variable, which is shown in Monte Carlo simulations to possess desirable small sample properties. The models and estimators are applied to data on US patents and R&D expenditure.

Key Words: dynamic count panel data, individual effects, predetermined regressors, Generalised Method of Moments, pre-sample information

JEL Classification: C23, C25, O30

---

1. Introduction

Count data processes are commonplace in empirical microeconomics. In panel data applications of count models it will often be the case that the process under study is inherently dynamic so that the history of the count process itself becomes an important determinant of current outcomes. As in other panel data applications it is also likely that unobserved heterogeneity will be present, inducing persistently different counts across individuals or firms. This paper makes two contributions to the development of a modeling framework that allows for dynamic feedback and unobserved heterogeneity. First, the use of pre-sample information is investigated as a method of improving the efficiency of estimators in panel count models with predetermined covariates. Second, a linear feedback model for dynamic panel data applications is proposed. This modeling approach is then applied to the analysis of the relationship between R&D and patents for a panel of US firms.

In count data models, where a natural non-linearity is produced by the non-negative discrete nature of the data, standard Generalised Method of Moments (GMM) methods for the estimation of dynamic fixed effects models[1] are not directly applicable. The standard panel data estimator for count models with correlated fixed effects is the Poisson conditional maximum likelihood estimator proposed by Hausman, Hall and Griliches (1984). We show that this estimator is the same as the Poisson maximum likelihood estimator in a model with individual specific constants, as the Poisson model does not suffer from the incidental parameter problem. This estimator is equivalent to a moment estimator in a model where the fixed effects are replaced by the ratio of within group means. It is therefore inconsistent for the parameters of interest if the regressors are predetermined and therefore are not strictly exogenous.

For the case of predetermined regressors, Chamberlain (1992, 1993) and Wooldridge (1997) have developed a consistent quasi-differenced GMM estimator. This has been

---

[1]See Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991) and Ahn and Schmidt (1995), for example.

applied to the estimation of count data models with predetermined regressors by Montalvo (1997), Crépon and Duguet (1997), and Cincera (1997). In this paper we extend this quasi-differenced GMM estimator to the dynamic linear feedback model, and investigate its small sample behaviour by means of a Monte Carlo study.

Our investigations show that the quasi-differenced GMM estimator can be severely biased in small samples when regressors are highly persistent, and the instruments therefore weak predictors of the endogenous variables in the differenced model. As an alternative estimator we propose a pre-sample mean (PSM) estimator that replaces the fixed effect by the *pre-sample* mean of the dependent variable. This estimator is specifically designed to exploit data sets where information on the dependent variable is available for earlier years than information on the regressors. This is a common situation where administrative data, such as patents or innovation series, is matched with survey or accounting data.[2] This estimator is consistent when the number of pre-sample periods gets large.

To explicitly model the dynamics of the count process in panel data we consider the linear feedback model (LFM). This has its foundations in the Integer Valued Autoregressive (INAR) generalisation of the Poisson model to the ARMA process developed by Al-Osh and Alzaid (1987), McKenzie (1988) and Brännäs (1994). In this model the conditional mean of the count variable is modelled linearly in the history of the process. This specification is shown to be well adapted to economic applications and especially convenient for understanding the dynamic properties of count data processes.

An application is made to the data set of US firms' R&D expenditure and patenting activity used in Hall, Griliches and Hausman (1986) and related work. The issues addressed in this earlier work relate to quantifying the R&D cost of a patent and assessing the timing of returns to R&D. This is made difficult by the fact that not

---

[2]Examples of these types of data sets include the SPRU data in the UK, see Pavitt, Robson and Townsend (1987), and patent and R&D series in many countries.

only may R&D expenditures lead to patents but also patents may themselves induce future R&D expenditures. Consequently, R&D is unlikely to be strictly exogenous.

The layout of the paper is as follows. In the next section we discuss the Poisson maximum likelihood estimator for count data in a model with individual specific constants and show its equivalence with the Poisson conditional maximum likelihood estimator. The use of a quasi-differenced GMM estimator for a model with predetermined regressors is considered, and the pre-sample mean estimator is introduced. In section 3, the linear feedback model is presented and the estimation of the model is discussed. Section 4 presents results of a Monte Carlo simulation study and in section 5 the dynamic model and various estimators are applied to the patent-R&D data.

2. Fixed Effects Models for Count Data

Let y_{it} denote the discrete count variable to be explained for subject i, $i = 1, ..., N$, at time t, $t = 1, ..., T$; and let x_{it} denote a vector of explanatory variables. The exponential (or log-link) model of the form

$$E\left(y_{it}|x_{it}\right) = \exp\left(x'_{it}\beta\right), \qquad (2.1)$$

is commonly used for count data.[3] An important feature in panel data applications is unobserved heterogeneity or individual fixed effects. For count data models these effects are generally modelled multiplicatively as

$$\begin{aligned} E\left(y_{it}|x_{it}, \eta_i\right) &= \exp\left(x'_{it}\beta + \eta_i\right) \qquad (2.2) \\ &= \mu_{it}\nu_i, \end{aligned}$$

where $\mu_{it} \equiv \exp\left(x'_{it}\beta\right)$, and $\nu_i \equiv \exp\left(\eta_i\right)$ is a permanent scaling factor for the individual specific mean. This implicitly defines a regression model

$$y_{it} = \mu_{it}\nu_i + u_{it}. \qquad (2.3)$$

---

[3]See Gourieroux, Monfort and Trognon (1984), McCullagh and Nelder (1989), Cameron and Trivedi (1986) and Winkelmann (1997) for good general discussions of these models.

In general, it is likely that the unobserved fixed components η_i are correlated with the explanatory variables, $E(x_{it}\eta_i) \neq 0$, and therefore standard random effects estimators will be inconsistent.

2.1. Strictly Exogenous Regressors

When the x_{it} are *strictly exogenous,* the conditional mean of y_{it} satisfies

$$E(y_{it}|\nu_i, x_{it}) = E(y_{it}|\nu_i, x_{i1}, ..., x_{iT}). \tag{2.4}$$

For this case, several estimators for β have been proposed that allow for individual effects to be correlated with the regressors. Hausman, Hall and Griliches (1984) use the Poisson conditional maximum likelihood estimator (CMLE), conditioning on $\sum_{t=1}^{T} y_{it}$, which is the sufficient statistic for η_i. This method mimics the fixed effect logit approach of Chamberlain (1984). However, the Poisson maximum likelihood estimator (MLE) for β in a model with separate individual specific constants does not suffer from the incidental parameters problem, and is therefore consistent (and, we show below, is the same as the CMLE). To see this, note that the maximum likelihood first order conditions for the ν_i are given by

$$\frac{\partial \ln L}{\partial \nu_i} = \sum_{t=1}^{T} \frac{\partial(y_{it}\ln(\mu_{it}\nu_i) - \mu_{it}\nu_i)}{\partial \nu_i} = \sum_{t=1}^{T}\left(\frac{y_{it}}{\nu_i} - \mu_{it}\right) = 0$$

and therefore the MLE for ν_i is given by

$$\widehat{\nu}_{i(ML)} = \frac{\overline{y}_i}{\overline{\mu}_i},$$

where $\overline{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$ and $\overline{\mu}_i = T^{-1}\sum_{t=1}^{T} \exp(x'_{it}\beta)$. The MLE of the fixed effect is independent of ν_i. Substituting the fixed effects estimates in the first order conditions for β results in

$$\frac{\partial \ln L}{\partial \beta}(\widehat{\nu}_i) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mu_{it}\frac{\overline{y}_i}{\overline{\mu}_i}\right)x_{it} = 0.$$

When x_{it} is strictly exogenous,

$$\text{plim}_{N\to\infty}\frac{1}{N}\frac{\partial \ln L}{\partial \beta}(\widehat{\nu}_i) = \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{it} - \frac{\mu_{it}}{\overline{\mu}_i}\overline{u}_i\right)x_{it} = 0,$$

4

with $\overline{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$, and therefore the MLE for β is consistent.[4] It is further identical to the CMLE. The latter can be seen as follows. The Poisson conditional log likelihood function is given by[5]

$$\ln CL = \sum_{i=1}^{N} \sum_{t=1}^{T} \Gamma(y_{it} + 1) - \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \ln \left[\sum_{s=1}^{T} \exp\left(-(x_{it} - x_{is})'\beta\right) \right],$$

where $\Gamma(.)$ is the gamma function. The first-order condition for β is

$$
\begin{aligned}
\frac{\partial \ln CL}{\partial \beta} &= \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{y_{it}}{\sum_{s=1}^{T} \exp\left(-(x_{it} - x_{is})'\beta\right)} \sum_{s=1}^{T} \exp\left(-(x_{it} - x_{is})'\beta\right)(x_{it} - x_{is}) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} x_{it} - \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \frac{\sum_{s=1}^{T} x_{is} \exp\left(x_{is}'\beta\right)}{\sum_{s=1}^{T} \exp\left(x_{is}'\beta\right)} \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \left(y_{it} - \mu_{it} \frac{\overline{y}_i}{\overline{\mu}_i} \right),
\end{aligned}
$$

which is exactly the same as the MLE first order condition for β.

The first order conditions imply that the Poisson MLE for β is equivalent to the moment estimator in a model where the ratio of individual, or within group, means are used to approximate the individual specific effects. This *mean scaling* model is given by

$$y_{it} = \mu_{it} \frac{\overline{y}_i}{\overline{\mu}_i} + u_{it}^*, \tag{2.5}$$

where $u_{it}^* = u_{it} - \frac{\mu_{it}}{\overline{\mu}_i} \overline{u}_i$. We call this estimator the *within group mean scaling* estimator.[6] The results are summarised in the following lemma.

Lemma 1. *The Poisson MLE for β in the model with separate individual constants is consistent and the same as the Poisson CMLE. The estimator is identical to the moment estimator in a mean scaling model solving the moment conditions*

$$\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \left(y_{it} - \mu_{it} \frac{\overline{y}_i}{\overline{\mu}_i} \right) = 0, \tag{2.6}$$

where $\mu_{it} = \exp\left(x_{it}'\beta\right)$.

---

[4]Lancaster (1997) finds the same result for the Poisson model by means of a decomposition of the likelihood.

[5]See Hausman, Hall and Griliches (1984, p. 919).

[6]Clearly, the Poisson pseudo-likelihood results are preserved.

2.2. Predetermined Regressors

In economic applications it is commonly the case that there is feedback from the dependent variable to the explanatory variables. A regressor is predetermined when it is not correlated with current and future shocks, but it is correlated with past shocks:

$$E\left(x_{it}u_{it+j}\right) = 0, \quad j \geq 0$$
$$E\left(x_{it}u_{it-s}\right) \neq 0, \quad s \geq 1.$$

With predetermined regressors, condition (2.4) does not hold and therefore the estimator of β that solves the sample moment conditions (2.6) is no longer consistent since x_{it} is now correlated with u_{it}^* through $\overline{u}_i$. This result is analogous to the inconsistency result for the within group estimator for linear panel data models with predetermined regressors in short panels.[7]

Chamberlain (1992) and Wooldridge (1997) have proposed transformations that eliminate the fixed effect from the multiplicative model and generate orthogonality conditions that can be used for consistent estimation in count data models with predetermined regressors.[8] The transformation is

$$s_{it} = y_{it}\frac{\mu_{it-1}}{\mu_{it}} - y_{it-1} \tag{2.7}$$

$$= u_{it}\frac{\mu_{it-1}}{\mu_{it}} - u_{it-1}.$$

Let $x_i^{t-1} = (x_{i1}, ..., x_{it-1})$. When x_{it} is predetermined, the following $T\left(T-1\right)/2$ moment conditions hold:

$$E\left(s_{it}|x_i^{t-1}\right) = E_{\nu|x^{t-1}}\left(E\left(s_{it}|\nu_i, x_i^{t-1}\right)\right) = 0, \tag{2.8}$$

using the law of iterated expectations.

---

[7] See Nickell (1981).

[8] For quasi-differenced moment conditions that are valid when x_{it} is endogenously determined, see Wooldridge (1991) and Windmeijer (2000).

Orthogonality condition (2.8) can be used to consistently estimate the model parameters β by the GMM estimation technique.[9] This estimator minimises

$$\left(\frac{1}{N} \sum_{i=1}^{N} s_i' Z_i \right) W_N^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' s_i \right),$$

where s_i is the $T-1$ vector of residuals $\{s_{it}\}$, Z_i is the matrix of instruments and W_N is a weight matrix. The optimal weight matrix is given by

$$W_N = \frac{1}{N} \sum_{i=1}^{N} Z_i' s_i(\widetilde{\beta}) s_i(\widetilde{\beta})' Z_i,$$

where $s_i(\widetilde{\beta})$ is based on an initial consistent estimate $\widetilde{\beta}$.

When only x_{it-1} is used as an instrument, the GMM estimator solves the sample moment conditions

$$\sum_{i=1}^{N} \sum_{t=2}^{T} x_{it-1} s_{it} = -\sum_{i=1}^{N} \sum_{t=2}^{T} x_{it-1} \left(y_{it-1} - \mu_{it-1} \frac{y_{it}}{\mu_{it}} \right) = 0, \tag{2.9}$$

which are similar to the first-order condition of the mean scaling estimator (2.6), but *leads* of y and μ are used to estimate ν_i instead of sample means.

2.3. Pre-Sample Mean Estimator

A common problem with the GMM estimation of (quasi-)differenced models using lags as instruments to predict future changes is that when economic series are highly persistent the instruments are very weak predictors. This results in a small sample bias of the GMM estimator in the direction of the estimator that uses the explanatory variables themselves, which are endogenous in the differenced equation, as instruments.[10]

In the patents and R&D application presented below, as in many applications, both series are highly persistent over time and there is pre-sample information (i.e. a longer time series) on the dependent variable (the patent series). An estimator that

---

[9] See Hansen (1982).
[10] See Staiger and Stock (1997), and Blundell and Bond (1998).

utilises this pre-sample information is a moment estimator in the model

$$y_{it} = \exp\left(\beta_0^* + x_{it}'\beta + \phi \ln \overline{y}_{ip}\right) + \varepsilon_{it},$$

where $\overline{y}_{ip} = \frac{1}{TP} \sum_{r=0}^{TP-1} y_{i0-r}$ is the pre-sample mean of y; TP is the number of pre-sample observations and ϕ is a parameter to be estimated. For convenience, we assume that there is a single regressor in the model, the extension to multiple regressors is straightforward. The moment estimator solves the just identified sample moment conditions

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \left(y_{it} - \exp\left(\beta_0^* + x_{it}\beta + \phi \ln \overline{y}_{ip}\right)\right) = 0, \tag{2.10}$$

where $z_{it} = \left(1, x_{it}, \ln \overline{y}_{ip}\right)$. We call this estimator the *pre-sample mean estimator*. The following lemma and corollary state the general conditions under which this estimator is consistent.

Lemma 2. *The moment estimator that solves the sample moment conditions (2.10) is consistent for $N \to \infty$ and $TP \to \infty$, if*

$$\ln\left(\overline{y}_{ip}\right) = \ln\left(\frac{1}{TP} \sum_{r=0}^{TP-1} y_{i0-r}\right) = a + b\eta_i + o_{p(TP)}\left(1\right), \tag{2.11}$$

where a random variable q_{TP} is $o_{p(TP)}\left(1\right)$ if $\text{plim}_{TP\to\infty}\left(q_{TP}\right) = 0$. The parameters are then given by $\phi = \frac{1}{b}$ and $\beta_0^* = \beta_0 - \frac{a}{b}$.

Proof: Using property (2.11), the moments conditions (2.10) can be written as

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \left(\exp\left(\beta_0 + x_{it}\beta + \eta_i\right) - \exp\left(\beta_0^* + x_{it}\beta + \phi\left(a + b\eta_i + o_{p(TP)}\left(1\right)\right)\right) + u_{it}\right) = 0.$$

The corresponding population moment conditions hold for $\phi = \frac{1}{b}$ and $\beta_0^* = \beta_0 - \frac{a}{b}$ as in that case

$$E_{TP\to\infty} \sum_{t=1}^{T} z_{it} \left(\exp\left(\beta_0 + x_{it}\beta + \eta_i\right)\left(1 - \exp\left(o_{p(TP)}\left(1\right)\right)\right) + u_{it}\right) = 0,$$

as the term involving $\left(1 - \exp\left(o_{p(TP)}\left(1\right)\right)\right)$ vanishes when $TP \to \infty$.

The next corollary states the conditions that the x_{it} process has to satisfy in order for the conditions of Lemma 2 to hold.

8

Corollary 3. *The condition (2.11) holds and the pre-sample mean estimator is consistent for $N \to \infty$ and $TP \to \infty$ if the x_{it} process is a stationary i.i.d. process with finite moments and its (long-run) mean proportional to η_i.*

Proof: Write the x_{it} process that satisfies the conditions of the corollary as

$$x_{it} = \lambda\eta_i + v_{it}.$$

Using a Taylor series expansion, the expectation of y_{it} conditional on η_i is then given by

$$
\begin{aligned}
E\left(y_{it}|\eta_i\right) &= E\left(\exp\left(\beta_0 + x_{it}\beta + \eta_i\right) + u_{it}\right) \\
&= E\left(\exp\left(\beta_0 + (1 + \beta\lambda)\eta_i\right)v_{it}^*\right),
\end{aligned}
$$

where $v_{it}^* = \sum_{j=0}^{\infty}\frac{\beta^j}{j!}v_{it}^j$. Therefore, by the law of large numbers, the logarithm of the pre-sample mean is equal to

$$\ln\left(\overline{y}_{ip}\right) = a + (1 + \beta\lambda)\eta_i + o_{p(TP)}(1),$$

where $a = \beta_0 + \ln\left(\sum_{j=0}^{\infty}\frac{\beta^j}{j!}E\left(v^j\right)\right)$.

An example of a stationary x_{it} process that satisfies the conditions of the corollary is $x_{it} = \rho x_{it-1} + \tau\eta_i + \varepsilon_{it}$, with the ε_{it} i.i.d with finite moments. Then $\lambda = \frac{\tau}{1-\rho}$ and $v_{it} = \sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j}$. Note that if x_{it} is further normally distributed, a is given by $\beta_0 + \frac{1}{2}\beta^2\sigma_v^2$.

3. Linear Feedback Model

Introducing dynamics into models of the form (2.2) is not straightforward since the conditional mean is required to remain positive. Inclusion of functions of the lagged dependent variable in the exponential function can lead to explosive series or to problems with transforming zero values. The dynamic specification considered here

is a linear feedback model. The conditional mean in the linear feedback model (LFM) is defined as

$$
\begin{aligned}
E\left(y_{it} | y_{it-1}, x_{it}, \nu_i\right) &= \gamma y_{it-1} + \exp(x_{it}'\beta)\nu_i \quad\quad (3.1) \\
&\equiv \gamma y_{it-1} + \mu_{it}\nu_i.
\end{aligned}
$$

The LFM has its origins in the Integer-Valued Autoregressive (INAR) process,[11] and can be motivated as an entry-exit process with the probability of exit equal to $(1-\gamma)$. The correlation over time in the INAR(1) model without additional regressors is given by

$$
corr(y_{it}, y_{it-j}) = \gamma^j,
$$

with $\gamma > 0$, as only positive association is possible. Since $\mu_{it}\nu_i$ is non-negative, the mean value for y_{it} is bounded below by γy_{it-1}.

Even when the x_{it} are strictly exogenous, the within groups mean scaling estimator will be inconsistent for small T, as the lagged dependent variable is a predetermined variable.[12] For estimation by GMM, the quasi-differenced transformation for the LFM model is given by

$$
s_{it} = (y_{it} - \gamma y_{it-1})\frac{\mu_{it-1}}{\mu_{it}} - (y_{it-1} - \gamma y_{it-2}) = u_{it}\frac{\mu_{it-1}}{\mu_{it}} - u_{it-1}, \quad\quad (3.2)
$$

where $u_{it} = y_{it} - E\left(y_{it} | y_{it-1}, x_{it}, \nu_i\right)$. In the case of a single regressor x_{it}, which is strictly exogenous, the following $(T-1)(T-2)/2 + T(T-2)$ moment conditions hold

$$
E(s_{it} | y_{i1}, ..., y_{it-2}, x_{i1}, ..., x_{iT}) = 0. \quad\quad (3.3)
$$

For predetermined x_{it} the $(T-1)(T-2)/2 + T(T-1)/2 - 1$ moment conditions are

$$
E(s_{it} | y_{i1}, ..., y_{it-2}, x_{i1}, ..., x_{it-1}) = 0. \quad\quad (3.4)
$$

---

[11]See Al-Osh and Alzaid (1987), McKenzie (1988), Ronning and Jung (1992), Brännäs (1994) and Böckenholt (1999). For extension of the INAR(1) to the INAR(p) model see Alzaid and Al-Osh (1990) and Jin-Guan and Yuan (1991).

[12]Note that for the LFM the within groups mean scaling estimator is no longer equivalent to the Poisson fixed effects MLE.

The just identified sample moments for the pre-sample mean estimator for the LFM are given by

$$\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}\left(y_{it} - \gamma y_{it-1} - \exp\left(\beta_0^* + x_{it}\beta + \phi\ln\overline{y}_{ip}\right)\right) = 0, \tag{3.5}$$

where $z_{it} = \left(1, y_{it-1}, x_{it}, \ln\overline{y}_{ip}\right)$. This estimator is consistent under the same conditions as stated in Lemma 2 and Corollary 3. The constant β_0^* is in this case further shifted by $\ln(1-\gamma)/\phi$, as the (long-run) mean of y_{it} conditional on η_i is given by

$$
\begin{aligned}
E\left(y_{it}|\eta_i\right) &= E\left(\sum_{j=0}^{\infty}\gamma^j \exp\left(\beta_0 + x_{it-j}\beta + \eta_i\right)\right) \\
&= E\left(\sum_{j=0}^{\infty}\gamma^j \exp\left(\beta_0 + (1+\lambda\beta)\eta_i + \beta v_{it-j}\right)\right) \\
&= \frac{1}{1-\gamma}\left(\exp\left(\beta_0 + (1+\beta\lambda)\eta_i\right)\sum_{j=0}^{\infty}\frac{\beta^j}{j!}E\left(v^j\right)\right).
\end{aligned}
$$

4. Monte Carlo Results

Before turning to the empirical application it is important to provide an evaluation of the performance of the estimators described in the previous sections for the linear feedback model. First, results of some Monte Carlo experiments are presented with a design where the explanatory variables are not very persistent, and so the quasi-differenced GMM estimator is expected to perform relatively well. The performances of the estimators are then investigated in the linear feedback model for x_{it} processes that are more persistent over time. The pre-sample mean estimator is found to have smaller bias and lower root mean squared error in almost all settings.

The data generating processes are

$$
\begin{aligned}
y_{it} &\sim Poisson\left(\gamma y_{it-1} + \exp\left(\beta x_{it} + \eta_i\right)\right) \\
x_{it} &= \rho x_{it-1} + \tau\eta_i + \varepsilon_{it} \\
x_{i0} &= \frac{\tau}{1-\rho}\eta_i + \xi_i
\end{aligned}
$$

11

$$y_{i0} \sim Poisson\left(\exp\left(x_{i0}\beta + \eta_i\right)\right)$$

$$\eta_i \sim N\left[0, \sigma_\eta^2\right]$$

$$\varepsilon_{it} \sim N\left[0, \sigma_\varepsilon^2\right]$$

$$\xi_i \sim N\left[0, \frac{\sigma_\varepsilon^2}{1-\rho^2}\right].$$

Fifty periods are generated before the sample is drawn.

Results for the data generating process with parameter values $\gamma = 0.5$, $\beta = 0.5$, $\rho = 0.5$, $\tau = 0.1$, $\sigma_\eta^2 = 0.5$ and $\sigma_\varepsilon^2 = 0.5$ are presented in Tables 4.1 and 4.2 for $T = 4$ and $T = 8$ respectively and sample sizes $N = 100$, $N = 500$, and $N = 1000$. The bias and root mean squared error (RMSE) are reported for various estimators of γ and β, for 1000 replications.

The tables report results for a levels estimator that does not take account of unobserved heterogeneity and solves the moment conditions

$$\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}\left(y_{it} - \gamma y_{it-1} - \exp\left(\beta_0 + x_{it}\beta\right)\right) = 0, \tag{4.1}$$

with $z_{it} = (1, y_{it-1}, x_{it})$. These estimates are denoted γlev and βlev. As is clear from the tables, these level estimates are considerably biased upwards.

Results for the within group mean scaling estimator are also reported. This estimator solves the moment conditions

$$\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}\left(y_{it} - \gamma y_{it-1} - \mu_{it}\frac{\overline{y}_i - \gamma\overline{y}_{i,-1}}{\overline{\mu}_i}\right) = 0, \tag{4.2}$$

with $z_{it} = (y_{it-1}, x_{it})$, $\overline{y}_i = \frac{1}{T-1}\sum_{t=2}^{T} y_{it}$, $\overline{y}_{i,-1} = \frac{1}{T-1}\sum_{t=2}^{T} y_{it-1}$, and $\overline{\mu}_i = \frac{1}{T-1}\sum_{t=2}^{T}\exp\left(x_{it}\beta\right)$. These estimates are denoted γwg and βwg in the tables. As expected, the within group estimates are biased downwards.

The pre-sample mean estimators that solve (3.5) are denoted γpsm and βpsm. Results are reported for 8, 25 and 50 pre-sample periods respectively. The bias and RMSE of these estimates are relatively small, and both decrease as the number of pre-sample periods increases. Increasing the sample size N and/or T does improve

12

the RMSE, but only marginally so.[13]

Table 4.1: Monte Carlo results for Linear Feedback Model, T=4

	$\gamma = 0.5; \beta = 0.5; \rho = 0.5; \tau = 0.1; \sigma_\eta^2 = 0.5; \sigma_\varepsilon^2 = 0.5$					
	$N = 100$		$N = 500$		$N = 1000$	
	bias	rmse	bias	rmse	bias	rmse
γlev	0.262	0.270	0.274	0.276	0.276	0.278
βlev	0.547	0.669	0.506	0.581	0.505	0.561
γwg	-0.450	0.461	-0.445	0.448	-0.446	0.447
βwg	-0.262	0.275	-0.261	0.263	-0.262	0.263
$\gamma psm8$	0.054	0.095	0.066	0.076	0.065	0.070
$\beta psm8$	0.058	0.145	0.063	0.087	0.059	0.072
$\gamma psm25$	0.024	0.076	0.030	0.046	0.031	0.039
$\beta psm25$	0.024	0.116	0.029	0.058	0.027	0.044
$\gamma psm50$	0.009	0.072	0.015	0.036	0.016	0.028
$\beta psm50$	0.009	0.107	0.015	0.050	0.013	0.035
$\gamma qdpr$	-0.236	0.311	-0.092	0.135	-0.060	0.094
$\beta qdpr$	-0.237	0.315	-0.122	0.184	-0.086	0.141
$\gamma qdse$	-0.236	0.295	-0.083	0.118	-0.050	0.082
$\beta qdse$	-0.180	0.222	-0.080	0.116	-0.050	0.090

Notes to Table: Number of replications is 1000, *lev* is levels without fixed effects, *wg* is within group mean scaling, *psmTP* is pre-sample mean with TP pre sample periods, *qdpr(se)* is quasi-differenced GMM assuming x_{it} is predetermined (strictly exogenous). The instrument set for the qd GMM estimator includes time dummies.

The final set of results are for the quasi-differenced GMM estimator. Assuming that the x_{it} are predetermined, and so only using moment conditions (3.4) results in the estimates $\gamma qdpr$ and $\beta qdpr$. Assuming that the regressor is strictly exogenous, and therefore utilising the larger set of moment conditions (3.3) results in the estimates $\gamma qdse$ and $\beta qdse$. Given the way the data is generated the strict exogeneity assumption

---

[13]Using both within and pre-sample information to construct the individual mean of the dependent variable improves the performance of the estimator for γ slightly, but the bias and RMSE of the estimator for β increase, especially for low values of T and TP.

Table 4.2: Monte Carlo results for Linear Feedback Model, T=8

	$\gamma = 0.5; \beta = 0.5; \rho = 0.5; \tau = 0.1; \sigma_\eta^2 = 0.5; \sigma_\varepsilon^2 = 0.5$					
	$N = 100$		$N = 500$		$N = 1000$	
	bias	rmse	bias	rmse	bias	rmse
γlev	0.263	0.268	0.274	0.275	0.278	0.279
βlev	0.504	0.617	0.505	0.565	0.554	0.561
γwg	-0.186	0.195	-0.184	0.186	-0.184	0.185
βwg	-0.125	0.140	-0.126	0.129	-0.128	0.129
$\gamma psm8$	0.074	0.089	0.084	0.087	0.087	0.088
$\beta psm8$	0.081	0.119	0.087	0.096	0.086	0.090
$\gamma psm25$	0.032	0.056	0.041	0.046	0.043	0.046
$\beta psm25$	0.035	0.082	0.039	0.051	0.039	0.045
$\gamma psm50$	0.015	0.047	0.023	0.032	0.025	0.029
$\beta psm50$	0.018	0.071	0.022	0.038	0.021	0.030
$\gamma qdpr$	-0.301	0.317	-0.097	0.106	-0.056	0.066
$\beta qdpr$	-0.278	0.290	-0.137	0.147	-0.091	0.102
$\gamma qdse$	-0.347	0.359	-0.105	0.113	-0.060	0.066
$\beta qdse$	-0.256	0.261	-0.112	0.118	-0.072	0.078

Notes to Table: Number of replications is 1000, lev is levels without fixed effects, wg is within group mean scaling, $psmTP$ is pre-sample mean with TP pre sample periods, $qdpr(se)$ is quasi-differenced GMM assuming x_{it} is predetermined (strictly exogenous). The instrument set for the qd GMM estimator includes time dummies.

is valid. For both values of T the quasi-differenced GMM estimator is downward biased for small values of N. The bias and RMSE decrease when the sample size N increases, but the results are very similar for $T = 4$ and $T = 8$. Using the extra moment conditions due to the strict exogeneity of the regressor improves the performance of the estimator of β but not of that of γ. Especially when $T = 8$, $\gamma qdse$ performs worse than $\gamma qdpr$, both in terms of bias and RMSE.

Comparing the GMM quasi-differenced estimates ($qdpr$, $qdse$) to the pre-sample mean estimates, it becomes clear that for this particular data generating process the pre-sample mean estimator outperforms the GMM quasi-differenced estimator in almost all cases, even for a number of pre-sample observations as small as 8. In particular, the quasi-differenced GMM estimator for the linear feedback model seems to suffer from a poor small sample performance.

The performance of the estimators in the linear feedback model for x_{it} processes that are more persistent over time are considered next. The parameters are set to $\alpha = 0.7$, $\beta = 1$, $\tau = 0$, so x_{it} is not correlated with η_i in this case, and $\sigma_\eta^2 = 0.5$. The sample size is $N = 500$, $T = 8$. Two different processes for the regressor are considered, a persistent process with $\rho = 0.9$ and $\sigma_\varepsilon^2 = 0.05$, and an even more persistent series with $\rho = 0.95$ and $\sigma_\varepsilon^2 = 0.015$. Bias and RMSE for the estimators for these two data generating processes are given in Table 4.3, again for 1000 replications.

When $\rho = 0.9$ and $\sigma_\varepsilon^2 = 0.05$, the results are very similar to those presented in Tables 4.1 and 4.2, with the pre-sample mean estimator outperforming the quasi-differenced GMM estimator. However, the bias and RMSE for the pre-sample mean estimator of β increase when the x_{it} series become very persistent, when $\rho = 0.95$ and $\sigma_\varepsilon^2 = 0.015$. This is due to the fact that the information contained in $\ln \overline{y}_{ip}$ becomes more and more like that contained in x_{it} when the x_{it} process becomes more persistent. Therefore, the separate estimation of β and the parameter on $\overline{y}_{ip}$ becomes problematic due to the multicollinearity. As the variance of the fixed effects η_i is relatively large, the estimation of the parameter on $\overline{y}_{ip}$ will get a larger weight, with the estimate for

15

β biased towards zero.

There is a very substantial downward bias when the parameters are estimated by quasi-differenced GMM, and the problem of instrumenting the differenced x_{it} series by lagged values becomes clear, as the RMSE of $\beta qdpr$ is large, much larger than those for the pre-sample mean estimator.

Table 4.3: Monte Carlo results for Linear Feedback Model, persistent x series

$\alpha = 0.7; \beta = 1; \tau = 0; \sigma_\eta^2 = 0.5, N = 500, T = 8$				
	$\rho = 0.90; \sigma_\varepsilon^2 = 0.05$		$\rho = 0.95; \sigma_\varepsilon^2 = 0.015$	
	bias	rmse	bias	rmse
γlev	0.181	0.182	0.183	0.184
βlev	0.433	0.481	0.249	0.316
γwg	-0.245	0.247	-0.271	0.272
βwg	-0.367	0.374	-0.361	0.380
$\gamma psm8$	0.094	0.097	0.081	0.084
$\beta psm8$	-0.099	0.146	-0.366	0.382
$\gamma psm25$	0.074	0.077	0.066	0.069
$\beta psm25$	-0.056	0.104	-0.286	0.302
$\gamma psm50$	0.056	0.060	0.055	0.059
$\beta psm50$	-0.028	0.080	-0.195	0.215
$\gamma qdpr$	-0.128	0.142	-0.127	0.146
$\beta qdpr$	-0.427	0.505	-0.435	0.714
$\gamma qdse$	-0.161	0.172	-0.160	0.175
$\beta qdse$	-0.344	0.363	-0.314	0.366

Notes to Table: Number of replications is 1000, *lev* is levels without fixed effects, *wg* is within group mean scaling, $psm(s)TP$ is pre-sample mean (scaling) with TP pre sample periods, $qdpr(se)$ is quasi-differenced GMM assuming x_{it} is predetermined (strictly exogenous). The instrument set for the q-d GMM estimators includes time dummies.

5. An Application to the R&D and Patents Relationship

The data used to examine the properties of the alternative dynamic specifications are the patent-R&D panel data of Hall, Griliches and Hausman (1986).[14] The data contain information on R&D expenditures of large US firms between 1972 and 1979 and a count of patents applied for and subsequently granted between 1965 and 1979. The firms in the data set are 407 firms for which there is a pre-sample history of the dependent variable.[15]

Some descriptive statistics and the first order autocorrelations of the lnR&D and patent series for the two groups are given in Table 5.1. Both series are highly autoregressive.

Before presenting the estimation results for the LFM specification, we give an economic interpretation of the linear feedback model in terms of the R&D-patents application. The relationship being estimated between patents and R&D can be interpreted as a knowledge production function describing the production of patents from current and past R&D investment. A simple way to write this relationship is

$$Q_{it} = g(R_{it}, R_{it-1}, ..., \beta, \nu_i) \tag{5.1}$$

where Q_{it} is some latent measure of technological output of firm i in period t, R_{it} measures the corresponding R&D investment, β represents the vector of unknown technology parameters and ν_i captures the firm specific propensity to patent. Observed patents, P_{it}, are a noisy indicator of a firm's technological output,

$$P_{it} = Q_{it} + \varepsilon_{it}, \tag{5.2}$$

with $E(\varepsilon_{it}|R_{it}, R_{it-1}, ..., \nu_i) = 0$.

---

[14]This data set is described in Bound et al (1982) and derivative data sets have been used in many applications including Hausman, Hall and Griliches (1984).

[15]Due to a large heterogeneity in estimated coefficients, we have excluded 37 firms from the "Motors/Aircraft" industry.

Table 5.1: The R&D and Patents Data

# firms	407	
	lnR&D	Patents
Mean	1.312	35.25
Std. Dev.	1.815	79.09
Between Std. Dev.	1.795	77.91
Within Std. Dev.	0.281	14.09
Median	1.267	6
Minimum	-3.849	0
Maximum	6.682	906
Proportion of zeros	-	0.154
First order autocorrelation	0.925	0.884
se	0.029	0.017

Notes to Table: The first order autocorrelation estimated by System GMM takes account of fixed effects using the stacked linear and differenced GMM estimator for the mean stationary linear dynamic error components model, see Blundell and Bond (1998), instruments used are lagged levels dated t-j,..,t-5 for the differenced equation and lagged differences dated t-j+1 for the levels equation, where j=2 for the R&D series, and j=3 for the patent series.

Suppose that historic R&D investments are combined through a Cobb-Douglas technology to produce knowledge stock. In this case (5.1) becomes

$$Q_{it} = R_{it}^{\beta_1} R_{it-1}^{\beta_2} R_{it-2}^{\beta_3} ... \nu_i. \tag{5.3}$$

This motivates the conditional mean specification in a multiplicative distributed lag model

$$E\left(y_{it}|x_{it}, x_{it-1}, \ldots, x_{it-p}, \nu_i\right) = \exp(x'_{it}\beta_1 + x'_{it-1}\beta_2 + ... + x'_{it-p}\beta_{p+1})\nu_i \tag{5.4}$$

where y_{it} corresponds to P_{it} and x_{it} to $\ln R_{it}$. This is the standard model used in the R&D-patents literature.[16]

---

[16]For an evaluation of this model see Blundell, Griffith and Van Reenen (1995), and Blundell, Griffith and Windmeijer (1999).

If there is limited history on R&D in the data or the R&D series are highly persistent so that the series of lagged values of R_{it-j} are collinear, the alternative dynamic specification of the linear feedback model can be attractive. To provide an economic motivation for the LFM specification (3.1), consider writing the relationship (5.1) between patents and R&D as

$$P_{it} = k \left(R_{it}^{\beta} + (1 - \delta) R_{it-1}^{\beta} + .. \right) \nu_i + \varepsilon_{it}, \qquad (5.5)$$

in which k is a positive constant and where R&D investment depreciates exponentially at rate δ. Setting $\beta < 1$ allows for a decreasing return to within period investment. The distributed lag term in brackets represents the process by which patents are produced from R&D inputs.[17] Ignoring any feedback from patents to R&D the long run steady state for firm i may be written

$$P_i = \frac{k}{\delta} R_i^{\beta} \nu_i, \qquad (5.6)$$

so that β may be interpreted as the long run elasticity. Inverting (5.5) we have

$$P_{it} = k R_{it}^{\beta} \nu_i + (1 - \delta) P_{it-1} + u_{it}, \qquad (5.7)$$

in which $E(u_{it}|R_{it}, P_{it-1}, \nu_i) = 0$. Depending on the serial correlation structure of ε_{it}, the process for u_{it} may display some autocorrelation. Equation (5.7) is equivalent to the LFM model given in (3.1) in which the autoregressive coefficient γ estimates the depreciation factor $(1 - \delta)$ and the long run elasticity is given by β.

Results for the Linear Feedback Model are presented in Table 5.2. The levels estimation procedure that does not take account of fixed effects results in coefficients on the lagged dependent variable and $\ln R\&D$ of 0.89 and 0.90 respectively. In contrast, the within groups mean scaling estimates are much lower, 0.41 and 0.34 respectively.

---

[17]For example, using a specific parameterisation of the CES production function where (5.1) becomes

$$Q_{it} = k(R_{it}^{\beta} + (1 - \delta) R_{it-1}^{\beta} + ...)^{\frac{\psi}{\beta}} \nu_i.$$

Setting the returns to scale parameter ψ equal to β and using (5.2) results in (5.5).

Table 5.2: Results for the Linear Feedback Model

	(1) Levels		(2) Within		(3) PSM		(4) QD(A)		(5) QD(B)	
Pat_{-1}	0.891	0.030	0.413	0.055	0.841	0.031	-0.192	0.047	-0.215	0.085
$\ln R\&D$	0.898	0.069	0.342	0.075	0.506	0.148	0.033	0.085	0.173	0.053
Sargan (p)							87.05	(0.003)		

Notes to Table: The model includes year dummies. Standard errors are shown in small font and allow for a general covariance matrix over firms and time. Levels' is the estimator solving sample moments (4.1) , not taking account of individual effects. 'Within' is within group mean scaling, solving (4.2). PSM is pre-sample mean, and QD is quasi-differenced GMM. The pre-sample mean of Pat uses the years 1965-71.. Instruments in column (4), QD(A), are $(1,y_{it-2},...,y_{i1},x_{it-1},...,x_{i1})$, in column (5), QD(B) they are (time dummies,$\Delta y_{it-1},\Delta x_{it}$).. Sargan is the standard χ^2 test for overidentifying restrictions

In this dynamic model with fixed effects we expect the levels estimates to be biased upwards, and the within groups estimates to be downward biased, as was confirmed by the results of the Monte Carlo simulations. Moving next to the results of the estimator that utilise the available pre-sample information on patents, we find that for the pre-sample mean estimator the estimated coefficients lie in-between the levels and within estimates, with the coefficient on the lagged dependent variable equal to 0.84 and that on $\ln R\&D$ equal to 0.51. All these results, in terms of the direction of the biases, are very similar to the Monte Carlo results with persistent x_{it} series, with $\rho = 0.9$, as displayed in Table 4.3.

The results of the PSM estimator have a sensible economic interpretation. They imply that the depreciation rate of patents is approximately 15% and indicate that the short run elasticity of patents with respect to R&D is approximately $(1 - \gamma)\beta = 0.08$, whereas the long run elasticity, ignoring feedback, is approximately 0.5. This implies a much slower moving adjustment process than that implied by the results

of a multiplicative distributed lag model that have previously been reported in the literature.

The results of the quasi-differenced GMM estimator as displayed in column (5)A, assuming that $\ln R\&D$ is predetermined, seem puzzling. The coefficient on the lagged dependent variable is *negative,* whereas the coefficient on $\ln R\&D$, although positive, is not significantly different from zero. The suspicion that these results arise due to a weak instruments problem as both the patents and the $\ln R\&D$ series are highly persistent, is confirmed by the results in column (5)B, where the Δy_{it-1} and Δx_{it} themselves are used as the only instruments. The estimates of the coefficient on lagged patents are almost identical, and the coefficient on $\ln R\&D$ is again very small, although larger than in column (4). This is clearly an indication that the instruments are (very) weakly correlated with the endogenous variables in the quasi-differenced model.[18]

6. Summary and Conclusions

We have shown that for standard count panel data models with correlated fixed effects and strictly exogenous regressors, the Poisson maximum likelihood estimator for the slope parameters in a model with individual constants is consistent, the same as the conditional maximum likelihood estimator, and equivalent to a within groups mean scaling estimator. This estimator is inconsistent when regressors are predetermined

This paper has proposed a pre-sample mean estimator for count panel data models. This estimator extends the set of panel data estimators to cover the case where there is additional historic information on the dependent variable. This is a common occurrence in many data sets and is exemplified in the patent count and R&D data for the US where recorded patent counts extend back over a long period before R&D

---

[18]Although the Sargan test indicates that the model is misspecified, other model specifications, for example including further lags of patents, or using different moment conditions that allow for possible endogeneity of $\ln R\&D$ and/or measurement error in patents do not alter the results. For subgroups of firms, the Sargan test does not reject instrument validity, whereas coefficient estimates do not change.

expenditure is available. The pre-sample estimator developed in the paper is shown to be an attractive alternative to standard estimators when there are correlated individual effects and predetermined regressors. It is shown to be consistent for large pre-sample size. In Monte-Carlo simulations it is shown to perform well in comparison to a quasi-differenced GMM estimator especially when the underlying processes are highly persistent.

The paper has developed the linear feedback framework for the case of dynamic count data models. A commonly used dynamic specification for count data is the multiplicative distributed lag model, where lags of the regressors enter the exponential mean function. In contrast, for the linear feedback model, the lagged dependent count variable enters the conditional mean specification linearly. The pre-sample estimator is then extended to cover the linear feedback model. In a Monte Carlo study it is also shown that, for reasonable pre-sample sizes, the estimator outperforms standard estimators in terms of bias and root mean squared error.

The model and estimators have been applied to a well known data set of US patents and R&D expenditure. The proposed pre-sample mean estimator performs well in comparison to alternatives that use the quasi-differencing approach. This is explained by an examination of the properties of the model and data in the light of the results from the Monte-Carlo simulation. The estimates of the linear feedback model estimated by the pre-sample mean estimator indicate that the adjustment process is much slower moving than has previously been found.

References

[1] Ahn, S. C. and P. Schmidt (1995) 'Efficient Estimation of Models for Dynamic Panel Data', *Journal of Econometrics*, 68, 5-28.

[2] Al-Osh, M.A., and A.A. Alzaid (1987) 'First-Order Integer Valued Autoregressive (INAR(1)) Process', *Journal of Time Series Analysis*, 8, 261-275.

[3] Alzaid, A.A., and M.A. Al-Osh (1990) 'Integer-Valued pth Order Autoregressive Structure (INAR(p)) Process', *Journal of Applied Probability*, 27, 314-324.

[4] Arellano, M., and S. Bond (1991) 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies*, 58, 277-98.

[5] Blundell, R., and S. Bond (1998) 'Initial Conditions and Moment Restrictions in Dynamic Panel Data Models', *Journal of Econometrics,* 87, 115-144.

[6] Blundell, R., R. Griffith, and J. van Reenen (1995) 'Dynamic Count Data Models of Technological Innovation', *The Economic Journal*, 105, 333-344.

[7] Blundell, R., R. Griffith, and F. Windmeijer (1999) 'Individual Effects and Dynamics in Count Data Models', Institute for Fiscal Studies Working Paper No. W99/3, http://www.ifs.org.uk/research/workingpapers/wp993.

[8] Bound, J. C. Cummins, Z. Griliches, B.H. Hall and A. Jaffe (1982) "Who does R&D and who patents?" in *R&D, Patents and Productivity*, Z. Griliches (ed.) Chicago: University of Chicago Press

[9] Böckenholt, U. (1999) 'Mixed INAR(1) Poisson Regression Models: Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data', *Journal of Econometrics*, 89, 317-338.

[10] Brännäs, K. (1994) 'Estimation and Testing in Integer Valued AR(1) Models', Umeå Economic Studies Paper No. 335, Department of Economics, University of Umeå, February.

[11] Cameron, A., and P. Trivedi (1986) 'Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests', *Journal of Applied Econometrics*, 1, 29-53.

[12] Chamberlain, G. (1984) 'Panel Data' in: Griliches, Z., and M. Intriligator (eds.), *Handbook of Econometrics*, North Holland, Amsterdam.

[13] Chamberlain, G. (1992) 'Comment: Sequential Moment Restrictions in Panel Data', *Journal of Business & Economic Statistics*, 10, 20-26.

[14] Chamberlain, G. (1993) 'Feedback in Panel Data Models', mimeo, *Econometric Society Summer Meeting*, Boston, June.

[15] Cincera, M. (1997) 'Patents, R&D, and Technological Spillovers at the Firm Level: Some Evidence from Econometric Count Models for Panel Data, *Journal of Applied Econometrics*, 12, 265-280.

[16] Crépon, B. and E. Duguet (1997) 'Estimating the Innovation Function from Patent Numbers: GMM on Count Panel Data, *Journal of Applied Econometrics*, 12, 243-263.

[17] Gourieroux, C., A. Monfort, and A. Trognon (1984) 'Pseudo Maximum Likelihood Methods: Applications to Poisson Models', *Econometrica*, 52, 701-720.

[18] Hall, B., Z. Griliches, and J. Hausman (1986) 'Patents and R and D: Is There a Lag?', *International Economic Review*, 27, 265-283.

[19] Hansen, L.P. (1982) 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica*, 50, 1029-1054.

[20] Hausman, J., B. Hall, and Z. Griliches (1984) 'Econometric Models for Count Data and an Application to the Patents-R&D Relationship' *Econometrica*, 52, 909-938.

[21] Holtz-Eakin, D., W. Newey, and H. Rosen (1988) 'Estimating Vector Autoregressions with Panel Data', *Econometrica,* 56, 1371-1395.

[22] Jin-Guan, D., and L. Yuan (1991) 'The Integer-Valued Autoregressive (INAR(p)) Model' *Journal of Time Series Analysis*, 12, 129-142.

[23] Lancaster, T. (1997), Orthogonal Parameters and Panel Data, mimeo, Brown University.

[24] McKenzie, E. (1988) 'Some ARMA Models for Dependent Sequences of Poisson Counts', *Advances in Applied Probability*, 20, 822-835.

[25] McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models,* Chapman and Hall, Monographs on Statistics and Applied Probability 37.

[26] Montalvo, J.G. (1997) 'GMM Estimation of Count-Panel-Data Models with Fixed Effects and Predetermined Instruments', *Journal of Business and Economic Statistics*, 15, 82-89.

[27] Nickell, S.J. (1981) 'Biases in Dynamic Models with Fixed Effects', *Econometrica*, 39, 359-382.

[28] Pavitt, K., M. Robson and J. Townsend (1987) 'The Size Distribution of Innovating Firms in the UK: 1945-83', *Journal of Industrial Economics*, 35, 297-316.

[29] Ronning, G., and R.C. Jung (1992) 'Estimation of a First Order Autoregressive Process with Poisson Marginals for Count Data', in: Fahrmeir, L., B. Francis, R. Gilchrist, and G. Tutz (eds.), 'Advances in GLIM and Statistical Modelling', *Lecture Notes in Statistics*, 78, 188-194, Springer-Verlag, Berlin.

[30] Staiger, D. and J.H. Stock (1997), Instrumental Variables Regression with Weak Instruments, *Econometrica, 65, 557-586.*

[31] Windmeijer, F. (2000), 'Moment Conditions for Fixed Effects Count Data Models with Endogenous Regressors', *Economics Letters* 68, 21-24.

[32] Winkelmann, R. (1997), *Econometric Analysis of Count Data, 2nd Edition*, Springer, Heidelberg.

[33] Wooldridge, J.M. (1991) 'Multiplicative Panel Data Models without the Strict Exogeneity Assumption', MIT Working Paper No. 574.

[34] Wooldridge, J.M. (1997) 'Multiplicative Panel Data Models without the Strict Exogeneity Assumption', *Econometric Theory*, 13, 667-678.