Individual effects and dynamics in count data models

Richard Blundell[a,b], Rachel Griffith[b], Frank Windmeijer[b,∗]

[a]*Department of Economics, University College London, Gower Street, London WC1E 6BT, UK*
[b]*Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK*

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. 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. © 2002 Elsevier Science B.V. All rights reserved.

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

---

∗ Corresponding author. Tel.: +44-20-7291-4800; fax: +44-20-7323-4780.
 E-mail address: f.windmeijer@ifs.org.uk (F. Windmeijer).

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 modelling 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 modelling 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 usual panel data estimator for count models with correlated fixed effects is the Poisson conditional maximum likelihood estimator proposed by Hausman et al. (1984). This estimator is the same as the Poisson maximum likelihood estimator in a model with individual specific constants. It 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 quasi-differenced GMM estimator that is consistent for such count data models. This has been 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 investigation shows that the quasi-differenced GMM estimator can be severely biased in small samples. This is particularly the case when regressors are highly persistent and the instruments are 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 consistent when the number of pre-sample periods is large and is shown in the Monte Carlo Study to have better finite sample properties than the quasi-differenced GMM estimator.

The PSM estimator we propose 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

---

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

administrative data is matched with survey or accounting data or where retrospective or recall data is collected. Examples of these types of data sets include: counts of patents or innovations matched to accounting data; [2] surveys such as the Health and Retirement Survey (HRS) which match individual administrative data to survey data; [3] long running longitudinal studies such as the panel study of income dynamics (PSID) which periodically collect additional data in intermittent years or extend the scope of questions. [4]

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 et al. (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 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. 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] Patent counts are available from the US patent office and extend back to 1965, these have been matched to accounting data in the US from 1972, and in the UK, also from 1972; see Hall et al. (1986) for the US and Bloom and Van Reenen (2000) for the UK. In the UK innovation counts had been collected by SPRU, the innovation count runs from 1945–1982 while the accounting data starts in 1972; see Pavitt et al. (1987) and Blundell et al. (1999).

[3] The HRS collects data on a sample of people of the age of 50 and over in four waves (biannually from 1992–1998), this has been matched to participants' social security records which extend back over their working lifetime; for example see Gustman et al. (1995).

[4] The PSID includes data on employment, housing, marital and fertility behaviour (among other things) from 1968–1997, wealth and pension data have been collected periodically, see for example Becketti et al. (1988) and there are currently plans to more detailed health and mortality data.

2. Fixed effects models for count data

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

$$\mathrm{E}(y_{it}|x_{it}) = \exp(x'_{it}\beta), \tag{2.1}$$

is commonly used for count data.[5] 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

$$\mathrm{E}(y_{it}|x_{it}, \eta_i) = \exp(x'_{it}\beta + \eta_i) = \mu_{it}v_i, \tag{2.2}$$

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

$$y_{it} = \mu_{it}v_i + u_{it}. \tag{2.3}$$

In general, it is likely that the unobserved fixed components η_i are correlated with the explanatory variables, $\mathrm{E}(x_{it}\eta_i) \neq 0$, 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

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

For this case, several estimators for β have been proposed that allow for individual effects to be correlated with the regressors. Hausman et al. (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 the same as the CMLE, see Blundell et al. (1999), and Lancaster (1999). The associated first order conditions imply that the Poisson MLE for β is equivalent to a moment estimator in a model where the ratio of individual, or within group, means are used to approximate the individual specific effects:

$$\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \left(y_{it} - \mu_{it} \frac{\bar{y}_i}{\bar{\mu}_i} \right) = 0 \tag{2.5}$$

---

[5] See Gourieroux et al. (1984), McCullagh and Nelder (1989), Cameron and Trivedi (1986, 1998) and Winkelmann (2000) for good general discussions of these models.

and the *mean scaling* model is given by

$$y_{it} = \mu_{it} \frac{\bar{y}_i}{\bar{\mu}_i} + u_{it}^*, \tag{2.6}$$

where $u_{it}^* = u_{it} - (\mu_{it}/\bar{\mu}_i)\bar{u}_i$. We call this estimator the *within group mean scaling* estimator.

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 pre-determined when it is not correlated with current and future shocks, but it is correlated with past shocks:

$$E(x_{it} u_{it+j}) = 0, \quad j \geqslant 0,$$

$$E(x_{it} u_{it-s}) \neq 0, \quad s \geqslant 1.$$

With predetermined regressors, condition (2.4) does not hold and therefore the estimator of β that solves the sample moment conditions (2.5) is no longer consistent since x_{it} is now correlated with u_{it}^* through $\bar{u}_i$. This result is analogous to the inconsistency result for the within group estimator for linear panel data models with predetermined regressors in short panels.[6]

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.[7] The transformation is

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

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

$$E(s_{it}|x_i^{t-1}) = E_{v|x^{t-1}}(E(s_{it}|v_i, x_i^{t-1})) = 0 \tag{2.8}$$

using the law of iterated expectations.

Orthogonality condition (2.8) can be used to consistently estimate the model parameters β by the GMM estimation technique.[8] This estimator minimises

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

---

[6] See Nickell (1981).

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

[8] See Hansen (1982).

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

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

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

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

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

which are similar to the first-order condition of the mean scaling estimator (2.5), but *leads* of y and μ are used to estimate v_i instead of sample means.

2.3. Pre-sample mean estimator

Common problems with the GMM estimation of (quasi-)differenced models are small sample bias and imprecision. This is particularly the case when economic series are highly persistent as the instruments are then weak predictors of future changes. 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. [9]

In the patents and R&D application presented below, as in many applications in economics, 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 utilises this pre-sample information is a moment estimator in the model

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

where $\bar{y}_{ip} = (1/TP)\sum_{r=0}^{TP-1} y_{i0-r}$ 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

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

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

---

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

Lemma 1. The moment estimator that solves the sample moment conditions (2.10) *is consistent for* $N \to \infty$ *and* $TP \to \infty$, *if*

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

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

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

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}\left(\exp(\beta_0 + x_{it}\beta + \eta_i)\right.$$

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

The corresponding population moment conditions hold for $\phi = 1/b$ and $\beta_0^* = \beta_0 - (a/b)$ as in that case

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

as the term involving $(1 - \exp(o_{p(TP)}(1)))$ 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 1 to hold.

Corollary 2. 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

$$E(y_{it}|\eta_i) = E(\exp(\beta_0 + x_{it}\beta + \eta_i) + u_{it})$$

$$= E(\exp(\beta_0 + (1 + \beta\lambda)\eta_i)v_{it}^*),$$

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

$$\ln(\bar{y}_{i_p}) = a + (1 + \beta\lambda)\eta_i + o_{p(TP)}(1),$$

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

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 $\lambda = \tau/(1 - \rho)$ 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

$$\mathrm{E}(y_{it}|y_{it-1}, x_{it}, v_i) = \gamma y_{it-1} + \exp(x_{it}'\beta)v_i$$

$$\equiv \gamma y_{it-1} + \mu_{it}v_i. \tag{3.1}$$

The LFM has its origins in the INAR process,[10] 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 $\mu_{it}v_i$ 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.[11] For estimation by GMM, the quasi-differenced transformation for the LFM model is given by

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

where $u_{it} = y_{it} - \mathrm{E}(y_{it}|y_{it-1}, x_{it}, v_i)$. For predetermined x_{it} the following $(T - 1)^2 - 1$ moment conditions hold

$$\mathrm{E}(s_{it}|y_{i1}, \ldots, y_{it-2}, x_{i1}, \ldots, x_{it-1}) = 0. \tag{3.3}$$

---

[10] 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).

[11] 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

$$\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}(y_{it} - \gamma y_{it-1} - \exp(\beta_0^* + x_{it}\beta + \phi \ln \bar{y}_{ip})) = 0, \tag{3.4}$$

where $z_{it} = (1, y_{it-1}, x_{it}, \ln \bar{y}_{ip})$. This estimator is consistent under the same conditions as stated in Lemma 1 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

$$E(y_{it}|\eta_i) = E\left(\sum_{j=0}^{\infty} \gamma^j \exp(\beta_0 + x_{it-j}\beta + \eta_i)\right)$$

$$= E\left(\sum_{j=0}^{\infty} \gamma^j \exp(\beta_0 + (1 + \lambda\beta)\eta_i + \beta v_{it-j})\right)$$

$$= \frac{1}{1 - \gamma}\left(\exp(\beta_0 + (1 + \beta\lambda)\eta_i)\sum_{j=0}^{\infty} \frac{\beta^j}{j!}E(v^j)\right).$$

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 than the GMM estimator in almost all settings, even with quite a small number of pre-sample periods.

The data generating processes are

$$y_{it} \sim Poisson(\gamma y_{it-1} + \exp(\beta x_{it} + \eta_i)),$$

$$x_{it} = \rho x_{it-1} + \tau \eta_i + \varepsilon_{it},$$

$$x_{i0} = \frac{\tau}{1 - \rho}\eta_i + \xi_i,$$

Table 1
Monte Carlo results for Linear Feedback Model, $T = 4$[a]

$\gamma = 0.5; \beta = 0.5; \rho = 0.5; \tau = 0.1; \sigma_\eta^2 = 0.5; \sigma_\varepsilon^2 = 0.5$							
		$N = 100$		$N = 500$		$N = 1000$	
		bias	rmse	bias	rmse	bias	rmse
LEVEL	γ	0.260	0.269	0.275	0.277	0.277	0.278
	β	0.565	0.698	0.506	0.585	0.505	0.567
WG	γ	−0.452	0.463	−0.445	0.447	−0.444	0.446
	β	−0.257	0.269	−0.260	0.263	−0.263	0.264
GMM	γ	−0.230	0.307	−0.090	0.131	−0.058	0.092
	β	−0.237	0.309	−0.126	0.184	−0.087	0.142
PSM	$\gamma(4)$	0.071	0.105	0.088	0.097	0.089	0.094
	$\beta(4)$	0.084	0.181	0.089	0.113	0.081	0.093
	$\gamma(8)$	0.051	0.089	0.068	0.077	0.068	0.073
	$\beta(8)$	0.061	0.148	0.067	0.090	0.059	0.072
	$\gamma(25)$	0.016	0.072	0.033	0.047	0.032	0.040
	$\beta(25)$	0.025	0.116	0.032	0.059	0.026	0.043
	$\gamma(50)$	0.007	0.070	0.017	0.037	0.016	0.028
	$\beta(50)$	0.016	0.108	0.018	0.050	0.012	0.035

[a] *Notes*: Number of replications is 1000, LEVEL is levels without fixed effects, WG is within group mean scaling, GMM is quasi-differenced GMM assuming x_{it} is predetermined and PSM is pre-sample mean with (TP) pre-sample periods. The instrument set for the GMM estimator includes time dummies.

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

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

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 1 and 2 for $T = 4$ and 8, respectively, and sample sizes $N = 100$, 500, and 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

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

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

Table 2
Monte Carlo results for Linear Feedback Model, $T = 8$[a]

$\gamma = 0.5; \beta = 0.5; \rho = 0.5; \tau = 0.1; \sigma_\eta^2 = 0.5; \sigma_\varepsilon^2 = 0.5$

		$N = 100$		$N = 500$		$N = 1000$	
		bias	rmse	bias	rmse	bias	rmse
LEVEL	γ	0.262	0.267	0.276	0.277	0.278	0.278
	β	0.504	0.599	0.504	0.568	0.555	0.562
WG	γ	−0.189	0.198	−0.184	0.186	−0.184	0.185
	β	−0.128	0.142	−0.126	0.129	−0.128	0.129
GMM	γ	−0.304	0.323	−0.096	0.106	−0.057	0.066
	β	−0.281	0.296	−0.136	0.147	−0.091	0.101
PSM	$\gamma(4)$	0.097	0.110	0.109	0.112	0.112	0.113
	$\beta(4)$	0.109	0.147	0.118	0.126	0.117	0.122
	$\gamma(8)$	0.073	0.088	0.084	0.087	0.086	0.087
	$\beta(8)$	0.080	0.118	0.086	0.094	0.085	0.089
	$\gamma(25)$	0.030	0.056	0.040	0.046	0.042	0.045
	$\beta(25)$	0.031	0.078	0.038	0.049	0.037	0.044
	$\gamma(50)$	0.013	0.048	0.022	0.031	0.024	0.028
	$\beta(50)$	0.015	0.067	0.021	0.036	0.020	0.029

[a]*Notes*: Number of replications is 1000, LEVEL is levels without fixed effects, WG is within group mean scaling, GMM is quasi-differenced GMM assuming x_{it} is predetermined and PSM is pre-sample mean with (TP) pre-sample periods. The instrument set for the GMM estimator includes time dummies.

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

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

with $z_{it} = (y_{it-1}, x_{it})$, $\bar{y}_i = 1/(T-1) \sum_{t=2}^{T} y_{it}$, $\bar{y}_{i,-1} = 1/(T-1) \sum_{t=2}^{T} y_{it-1}$, and $\bar{\mu}_i = 1/(T-1) \sum_{t=2}^{T} \exp(x_{it}\beta)$. As expected, the within group estimates are biased downwards.

Results for the quasi-differenced GMM estimator are reported, assuming that the x_{it} are predetermined, using moment conditions (3.3). 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 8.

Results for the pre-sample mean estimators that solve (3.4) are reported for 4, 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 the rmse, but only marginally so.[12]

Comparing the GMM quasi-differenced estimates 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 4. 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 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 1 and 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 $\ln \bar{y}_{ip}$ 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 $\bar{y}_{ip}$ becomes problematic due to the multicollinearity. As the variance of the fixed effects η_i is relatively large, the estimation of the parameter on $\bar{y}_{ip}$ will get a larger weight, with the estimate for 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 this estimator is large, much larger than that of the pre-sample mean estimator.

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 et al. (1986).[13] The data contain information on R&D expenditures of large US firms between 1972

---

[12] 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.

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

Table 3
Monte Carlo results for Linear Feedback Model, persistent x series[a]

$\alpha = 0.7; \beta = 1; \tau = 0; \sigma_\eta^2 = 0.5, N = 500, T = 8$					
		$\rho = 0.90; \sigma_\varepsilon^2 = 0.05$		$\rho = 0.95; \sigma_\varepsilon^2 = 0.015$	
		bias	rmse	bias	rmse
LEVEL	γ	0.181	0.182	0.183	0.184
	β	0.429	0.475	0.243	0.324
WG	γ	−0.244	0.246	−0.272	0.273
	β	−0.365	0.372	−0.357	0.376
GMM	γ	−0.127	0.144	−0.127	0.146
	β	−0.452	0.528	−0.458	0.738
PSM	$\gamma(4)$	0.097	0.106	0.089	0.091
	$\beta(4)$	−0.101	0.157	−0.372	0.390
	$\gamma(8)$	0.094	0.097	0.081	0.084
	$\beta(8)$	−0.097	0.148	−0.370	0.385
	$\gamma(25)$	0.073	0.077	0.067	0.070
	$\beta(25)$	−0.055	0.103	−0.288	0.304
	$\gamma(50)$	0.056	0.060	0.056	0.060
	$\beta(50)$	−0.026	0.080	−0.197	0.217

[a] *Notes*: Number of replications is 1000, LEVEL is levels without fixed effects, WG is within group mean scaling, GMM is quasi-differenced GMM assuming x_{it} is predetermined and PSM is pre-sample mean with (TP) pre-sample periods. The instrument set for the GMM estimator includes time dummies.

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. [14]

Some descriptive statistics and the first order autocorrelations of the ln R&D and patent series for the two groups are given in Table 4. 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

$$Q_{it} = g(R_{it}, R_{it-1}, \dots, \beta, v_i), \qquad (5.1)$$

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

Table 4
The R&D and patents data[a]

# firms	407	
	ln R&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

[a]*Notes*: 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 $t - j, \ldots, t - 5$ 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.

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 v_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 $\mathrm{E}(\varepsilon_{it}|R_{it}, R_{it-1}, \ldots, v_i) = 0$.

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

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

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

$$\mathrm{E}(y_{it}|x_{it}, x_{it-1}, \ldots, x_{it-p}, v_i) = \exp(x_{it}'\beta_1 + x_{it-1}'\beta_2 + \cdots + x_{it-p}'\beta_{p+1})v_i, \tag{5.4}$$

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. [15]

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,

---

[15] For an evaluation of this model see Blundell et al. (1995).

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

$$P_{it} = k(R_{it}^\beta + (1-\delta)R_{it-1}^\beta + \cdots)v_i + \varepsilon_{it}, \tag{5.5}$$

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. [16] Ignoring any feedback from patents to R&D the long run steady state for firm i may be written

$$P_i = \frac{k}{\delta}R_i^\beta v_i, \tag{5.6}$$

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

$$P_{it} = kR_{it}^\beta v_i + (1-\delta)P_{it-1} + u_{it}, \tag{5.7}$$

in which $E(u_{it}|R_{it}, P_{it-1}, v_i) = 0$. Depending on the serial correlation structure of ε_{it}, the process for u_{it} may display some autocorrelation. Eq. (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. 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. 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 utilises 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 3.

The results of the PSM estimator have a sensible economic interpretation. They imply that the depreciation rate of patents is approximately 15% and

---

[16] For example, using a specific parameterisation of the CES production function where (5.1) becomes $Q_{it} = k(R_{it}^\beta + (1-\delta)R_{it-1}^\beta + \cdots)^{\psi/\beta}v_i$. Setting the returns to scale parameter ψ equal to β and using (5.2) results in (5.5).

Table 5
Results for the Linear Feedback Model[a]

	(1) LEVEL		(2) WG		(3) PSM		(4) GMM(A)		(5) GMM(B)	
	Coeff	Se	Coeff	Se	Coeff	Se	Coeff	Se	Coeff	Se
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)		

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

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 [17] as displayed in column (4), 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), 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]

---

[17] The Gauss programme for the quasi-differenced GMM estimator is available at http://www.ifs.org.uk.
[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.

6. Summary and conclusions

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 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 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.

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 Monte-Carlo simulations it is shown to perform well in comparison to a quasi-differenced GMM estimator.

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.

Hall for providing the data on patents and R&D used in this study. The usual disclaimer applies.

References

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68, 5–28.

Al-Osh, M.A., Alzaid, A.A., 1987. First-order integer valued autoregressive (INAR(1)) process. Journal of Time Series Analysis 8, 261–275.

Alzaid, A.A., Al-Osh, M.A., 1990. Integer-valued *p*th order autoregressive structure (INAR(*p*)) process. Journal of Applied Probability 27, 314–324.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277–298.

Becketti, S., Gould, W., Lillard, L., Welch, F., 1988. The PSID after fourteen years: an evaluation. Journal of Labor Economics 6, 472–492.

Bloom, N., Van Reenen, J., 2000. Real options, patents, productivity and market value: evidence from a panel of British firms. Working paper series no. W00/21, Institute for Fiscal Studies, London.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–144.

Blundell, R., Griffith, R., Van Reenen, J., 1995. Dynamic count data models of technological innovation. The Economic Journal 105, 333–344.

Blundell, R., Griffith, R., Windmeijer, F., 1999. Individual effects and dynamics in count data models. Working paper series no. W99/3, Institute for Fiscal Studies.

Bound, J., Cummins, C., Griliches, Z., Hall, B.H., Jaffe, A., 1982. Who does R&D and who patents? In: Griliches, Z. (Ed.), R&D, Patents and Productivity. University of Chicago Press, Chicago.

Böckenholt, U., 1999. Mixed INAR(1) poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data. Journal of Econometrics 89, 317–338.

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.

Cameron, A., Trivedi, P., 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. Journal of Applied Econometrics 1, 29–53.

Cameron, A., Trivedi, P., 1998. Regression analysis of count data. Econometric Society Monograph, Cambridge University Press.

Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M. (Eds.), Handbook of Econometrics. North Holland, Amsterdam.

Chamberlain, G., 1992. Comment: sequential moment restrictions in panel data. Journal of Business & Economic Statistics 10, 20–26.

Chamberlain, G., 1993. Feedback in panel data models. Mimeo, Econometric Society Summer Meeting, Boston, June.

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.

Crépon, B., Duguet, E., 1997. Estimating the innovation function from patent numbers: GMM on count panel data. Journal of Applied Econometrics 12, 243–263.

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

Gustman, A.L, Mitchell, O.S., Steinmeier, T.L., 1995. Retirement research using the health and retirement survey. Journal of Human Resources 30, 557–583.

Hall, B., Griliches, Z., Hausman, J., 1986. Patents and R and D: is there a lag? International Economic Review 27, 265–283.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hausman, J., Hall, B., Griliches, G., 1984. Econometric models for count data and an application to the patents-R&D relationship. Econometrica 52, 909–938.

Holtz-Eakin, D., Newey, W., Rosen, H., 1988. Estimating vector autoregressions with panel data. Econometrica 56, 1371–1395.

Jin-Guan, D., Yuan, L., 1991. The integer-valued autoregressive (INAR(p)) model. Journal of Time Series Analysis 12, 129–142.

Lancaster, T., 1999. Orthogonal parameters and panel data. Mimeo, Brown University.

McKenzie, E., 1988. Some ARMA models for dependent sequences of poisson counts. Advances in Applied Probability 20, 822–835.

McCullagh, P., Nelder, J.A., 1989. Generalized linear models. Monographs on Statistics and Applied Probability, vol. 37. Chapman & Hall, London.

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.

Nickell, S.J., 1981. Biases in dynamic models with fixed effects. Econometrica 39, 359–382.

Pavitt, K., Robson, M., Townsend, J., 1987. The size distribution of innovating firms in the UK: 1945–83. Journal of Industrial Economics 35, 297–316.

Ronning, G., Jung, R.C., 1992. Estimation of a first order autoregressive process with Poisson marginals for count data. In: Fahrmeir, L., Francis, B., Gilchrist, R., Tutz, G. (Eds.), Advances in GLIM and Statistical Modelling, Lecture Notes in Statistics, vol. 78. Springer, Berlin, pp. 188–194.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Windmeijer, F., 2000. Moment conditions for fixed effects count data models with endogenous regressors. Economics Letters 68, 21–24.

Winkelmann, R., 2000. Econometric analysis of count data, 3rd Edition. Springer, Heidelberg.

Wooldridge, J.M., 1991. Multiplicative panel data models without the strict exogeneity assumption. MIT Working paper No. 574.

Wooldridge, J.M., 1997. Multiplicative panel data models without the strict exogeneity assumption. Econometric Theory 13, 667–678.