GMM Estimation of Count-Panel-Data Models With Fixed Effects and Predetermined Instruments

Jose G. MONTALVO
Department of Economics, Universitat Pompeu Fabra, Barcelona 08008, Spain, and IVIE, 46020 Valencia, Spain

The "traditional" approach to the estimation of count-panel-data models with fixed effects is the conditional maximum likelihood estimator. The pseudo maximum likelihood principle can be used in these models to obtain orthogonality conditions that generate a robust estimator. This estimator is inconsistent, however, when the instruments are not strictly exogenous. This article proposes a generalized method of moments estimator for count-panel-data models with fixed effects, based on a transformation of the conditional mean specification, that is consistent even when the explanatory variables are predetermined. Two applications are discussed, the relationship between patents and research and development expenditures and the explanation of technology transfer.

KEY WORDS: Conditional maximum likelihood estimator; Conditional moment restrictions; Patents–R&D relationship; Pseudo maximum likelihood principle; Technology transfer.

The use of count-data models to analyze economic dependent variables that take only nonnegative integer values has some tradition in econometrics. Count-data models are especially useful when a researcher wants to explain the number of times that an economic phenomenon takes place in a given period of time. Economic examples are the number of patents applied for by a firm in a particular year (Hall, Griliches, and Hausman 1986; Hall, Hausman, and Griliches 1984), the number of licensing agreements signed by a firm (Montalvo and Yafeh 1994), the number of cups of coffee that an individual consumes in a day (Mullahy 1986), or the number of medical consultations during a two-weeks period (Cameron and Trivedi 1986).

Regarding cross-section data, the basic framework for this kind of models has been the Poisson regression. The main disadvantage of this simple model is the assumption of equality between conditional mean and conditional variance. The usual finding when estimating a Poisson regression is overdispersion: The estimated variance is significantly larger than the estimated mean [for a recent treatment of the issue see, for instance, Efron (1992)]. Some authors have proposed generalizing the Poisson model using distributions derived from the Poisson that do not imply the equality between conditional expectation and variance. This is the case of the negative binomial of Hall et al. (1984) or the truncated Poisson of Mullahy (1986). In both cases the estimation procedure is maximum likelihood. The pseudo maximum likelihood (PML) methods proposed by Gourieroux, Monfort, and Trognon (1984a) can be used to obtain consistent estimators even when the family of probability distribution does not necessarily contain the true distribution. This approach was applied by Hall et al. (1986) and Cameron and Trivedi (1986).

With regard to count-panel-data models, the basic reference is the so-called conditional maximum likelihood (CML) fixed-effects specification used by Hall et al. (1984) [Gourieroux et al. (1984a) proposed double-indexed count data as an alternative to the specification of Hall et al. (1984) that can be estimated using a quasi-pseudo maximum likelihood estimator]. The fixed-effects model is appealing given that in many cases the individual effect has a significant correlation with the explanatory variables. For instance, firms that have a higher propensity to patent for unobserved reasons may invest more on research and development (R&D) because the returns of this kind of expenditures are higher than other investment projects. The main advantage of the CML specification is analytical tractability. The fact that the Poisson distribution belongs to the exponential family makes the sum of the dependent variable over time a Poisson distribution too. The problem with this approach is its dependence on the distributional assumptions. Additionally, it has a disadvantage: The consistency of the conditional maximum likelihood estimator (CMLE) relies on the strict exogeneity assumption. Wooldridge (1990) developed distribution-free estimation procedures for count-panel data and showed that consistency and asymptotic normality are guaranteed when the conditional mean is correctly specified. The need for strictly exogenous explanatory variables remains, however.

The strict exogeneity of the explanatory variables is a well-known requirement for consistency in the context of panel-data models. This assumption is likely to fail in many applications, however: For instance, firm patents are assumed to be a function of current and lag R&D expenditures. Because patents depend, most of the time, on additional R&D expenditures for their full development or improvement, R&D expenditures cannot be considered as strictly exogenous. In the case of licensing agreements and a firm's sales, it is obvious that the licenses obtained by a firm will generate higher sales in future periods.

Journal of Business & Economic Statistics
January 1997, Vol. 15, No. 1

In the linear case, any estimator that demeans or quasi-demeans the full sample in the original specification, such as the fixed or the random-effects estimators, will lead to inconsistent estimators when the explanatory variables are not strictly exogenous. Several alternatives have been proposed to estimate linear panel-data models without imposing the strict exogeneity assumption on the explanatory variables. Most of them are based on the first difference of the original specification because this transformation opens the possibility for finding valid instruments (see Anderson and Hsiao 1981; Holtz-Eakin, Newey, and Rosen 1988). Arellano and Bond (1991) proposed a forward demeaning procedure. Keane and Runkle (1992) suggested using a forward filtering based on Hayashy and Sims (1983) to eliminate serial correlation in the residuals. The set of valid instruments depends on the chosen procedure. These procedures, however, do not extend easily to the case of count models, given their nonlinear nature.

The purpose of this article is to present the theory and some economic applications of count-panel-data models when the strict exogeneity assumption is likely to fail. The outline of the article is as follows. Section 1 describes the CML approach to the estimation of count-panel-data models. Section 2 contains a discussion of several alternatives, proposed in the economic literature, to relax the distributional assumptions used in the CML approach. Section 3 presents a transformation strategy to obtain consistent estimators in the presence of explanatory variables that are not strictly exogenous. Section 4 contains two applications of the generalized method of moments (GMM) estimator described in Section 3. Section 4.1 describes an application to technology transfer to Japanese firms studied using cross-section techniques from Montalvo and Yafeh (1994). Section 4.2 presents an application to the relationship between patents and R&D, previously analyzed by Hall et al. (1984). In both cases there are good reasons to doubt the strict exogeneity of the explanatory variables, as we argued previously. Section 5 concludes.

1. THE CML APPROACH

Let y_{it} be a count variable, $y_{it} = 0, 1, 2, \ldots, \mu_i$ an unobservable random variable, and x_{it} a k-dimensional vector of conditioning variables. Assume that $y_{i1}, y_{i2}, \ldots, y_{it}$ are independent conditional on μ_i and x_i [$x_i \equiv (x_{i1}, x_{i2}, \ldots, x_{iT})$] and distributed as a Poisson distribution

$$P(y_{it}|x_i, \mu_i) = (\lambda_{it}^{y_{it}} \exp(-\lambda_{it}))/y_{it}!, \tag{1}$$

where the mean of the distribution is specified as

$$\lambda_{it} = \exp(\alpha_i + \beta_0' x_{it}) = \mu_i \exp(\beta_0' x_{it}). \tag{2}$$

This specification for the conditional mean has, by now, some tradition in economics (Hall et al. 1984; Hall et al. 1986). Equations (1) and (2) include the strict exogeneity of x_i conditional on μ_i.

Hall et al. (1984), following Andersen (1970), proposed a conditional maximum likelihood estimator to deal with the estimation of this model. The CMLE takes advantage of the fact that, given (1), the distribution of $n_i = \sum_{t=1}^{T} y_{it}$ is also a Poisson distribution with mean equal to $\sum_{t=1}^{T} \lambda_{it}$. The joint distribution of $y_i \equiv (y_{i1}, y_{i2}, \ldots, y_{iT})$ conditional on n_i, the unobservable effect, and the explanatory variables is

$$P(y_i|x_i, \alpha_i, n_i) = \frac{(\sum_{t=1}^{T} y_{it})!}{\prod_{t=1}^{T} y_{it}!} \prod_{t=1}^{T} p_{it}^{y_{it}}, \tag{3}$$

where

$$p_{it} \equiv \frac{\lambda_{it}}{\sum_{t=1}^{T} \lambda_{it}} = \frac{\exp(\beta_0' x_{it})}{\sum_{s=1}^{T} \exp(\beta_0' x_{is})} \tag{4}$$

and

$$\sum_{t=1}^{T} p_{it} = 1. \tag{5}$$

Equation (4) shows one of the main advantages of the CML approach under the Poisson specification: The conditioning eliminates the individual effects. Equations (3) and (5) show that the resulting conditional distribution is multinomial.

The log-likelihood function derived from the conditional approach is

$$L(\beta) = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \log(p_{it}(\beta)), \tag{6}$$

where the constant part has been eliminated for the sake of clarity.

2. PML ESTIMATION OF COUNT-PANEL DATA

To construct the likelihood function (6), it is necessary to assume that the probabilistic mechanism that generates the original observations is Poisson. There is often no theory available to justify the shape of the distribution, however. In many cases the only argument is the nonnegative integer nature of the dependent variable. Gourieroux et al. (1984a,b) showed how to obtain consistent and asymptotically normal estimators of parameters maximizing a likelihood function associated with a probability distribution that is not necessarily the true one. These procedures are grouped under the name of pseudo maximum likelihood (PML) methods [some other authors, like McCullagh and Nelder (1989), have used the terminology "quasi maximum likelihood methods"]. The general idea is to construct a function that has properties in common with the score of the original likelihood function. It is especially important that such a function have a zero expected value.

In particular, when the distribution belongs to the exponential family, Gourieroux et al. (1984a) have shown that the PML estimator (PMLE) is strongly consistent. This is the case of the Poisson and the multinomial distributions.

The main advantage of the PMLE is its robustness, in the sense that it is not necessary to specify a particular distribution function. It is usually sufficient to specify some characteristic features of the data. For instance, let us assume that y_{it} is a nonnegative random variable with conditional expectation specified as

$$E(y_{it}|x_i, \alpha_i) = \exp(\alpha_i + \beta_0' x_{it}), \tag{7}$$

an expression that coincides with the specification of the mean in Section 1 and keeps the property of representing nonnegative random variables. Given this conditional expectation, the function

$$\psi(y_i, x_i, \beta) = \sum_{t=1}^{T} y_{it} p_{it}(\beta)^{-1} \frac{\partial p_{it}(\beta)}{\partial \beta} \tag{8}$$

has an expected value, conditional on x_{it} and α_i, equal to 0:

$$\begin{aligned} E[\psi(y_i, x_i, \beta_0)|x_i, \alpha_i] &= \sum_{t=1}^{T} E(y_{it}|x_i, \alpha_i) p_{it}^{-1} \frac{\partial p_{it}(\beta_0)}{\partial \beta} \\ &= \sum_{t=1}^{T} \lambda_{it} \frac{\sum_{s=1}^{T} \lambda_{is}}{\lambda_{it}} \frac{\partial p_{it}(\beta_0)}{\partial \beta} \\ &= \sum_{s=1}^{T} \lambda_{is} \sum_{t=1}^{T} \frac{\partial p_{it}(\beta_0)}{\partial \beta} = 0 \end{aligned} \tag{9}$$

because $\sum_{t=1}^{T} [\partial p_{it}(\beta_0)]/\partial \beta = 0$ by (5). Under some regularity conditions (Hansen 1982) the law of large numbers will apply to the sample analog of (9),

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \psi(y_i, x_i, \beta) = 0. \tag{10}$$

Condition (10) can be used to construct a GMM estimator that is equivalent to the conditional MLE (in general, the MLE can be interpreted as a GMM estimator in which the moments are the elements of the score vector), given that the first-order conditions derived from (6) are

$$\sum_{n=1}^{N} \sum_{t=1}^{T} y_{it} p_{it}(\hat{\beta})^{-1} \frac{\partial p_{it}(\hat{\beta})}{\partial \beta} = 0. \tag{11}$$

This GMM estimator can be reinterpreted as a PMLE.

Wooldrige (1990) used the PML principle to propose, by analogy with the multinomial case, orthogonality conditions based on the function

$$u_{it}(\beta) = y_{it} - p_{it}(x_{it}, \beta) n_i \tag{12}$$

that have a conditional expected value equal to 0:

$$\begin{aligned} E[u_{it}(\beta_0)|x_i, \alpha_i] &= E[y_{it}|x_i, \alpha_i] - p_{it}(x_i, \beta_0) E[n_i|x_i, \alpha_i] \\ &= \lambda_{it} - \frac{\lambda_{it}}{\sum_{s=1}^{T} \lambda_{is}} \sum_{s=1}^{T} \lambda_{is} = 0. \end{aligned} \tag{13}$$

The implied orthogonality conditions take the form

$$E[Z_i(x_i)' u_i(\beta_0)] = 0, \tag{14}$$

where Z_i is any function of x_i that guarantees the existence of appropriate moments. Z_i could also be a function of β.

Under suitable regularity conditions (Hansen 1982), the limiting distribution of the estimator $\hat{\beta}$ that solves

$$\sum_{i=1}^{N} Z(x_i)' u_i(x_i, \hat{\beta}) = 0 \tag{15}$$

is given by

$$\sqrt{N}(\hat{\beta} - \beta_0) \sim N(0, \Lambda), \tag{16}$$

where

$$\Lambda \equiv (\Delta' \Phi^{-1} \Delta)^{-1}, \tag{17}$$

$$\Delta \equiv E[Z(x_i, \beta_0)' \nabla_\beta u_i(\beta_0)], \tag{18}$$

and

$$\Phi \equiv E[Z(x_i, \beta_0)' u_i(\beta_0) u_i(\beta_0)' Z(x_i, \beta_0)]. \tag{19}$$

Consistent estimators of these conditions can be obtained by using the following expressions:

$$\hat{\Delta} = \frac{1}{N} \sum_{i=1}^{N} Z(x_i, \hat{\beta})' \nabla_\beta u_i(\hat{\beta}) \tag{20}$$

and

$$\hat{\Phi} = \frac{1}{N} \sum_{i=1}^{N} Z(x_i, \hat{\beta})' u_i(\hat{\beta}) u_i(\hat{\beta})' Z(x_i, \hat{\beta}). \tag{21}$$

The CMLE is just a particular case of this class of estimators for a specific set of instruments. To see this, we can rewrite the score function (11) in functions of u_{it} as

$$\frac{\partial L(\beta)}{\partial \beta} = \sum_{n=1}^{N} \sum_{t=1}^{T} p_{it}(\beta)^{-1} \frac{\partial p_{it}(\beta)}{\partial \beta} u_{it} \tag{22}$$

given that

$$\sum_{t=1}^{T} p_{it}(\beta)^{-1} \frac{\partial p_{it}(\beta_0)}{\partial \beta} (p_{it}(\beta) n_i) = n_i \sum_{t=1}^{T} \frac{\partial p_{it}(\beta_0)}{\partial \beta} = 0. \tag{23}$$

3. COUNT–PANEL DATA WITH PREDETERMINED INSTRUMENTS

The consistency of the estimators derived in Sections 1 and 2 relies on the strict exogeneity of the instruments. In many relevant economic cases, however, the explanatory variables cannot be considered strictly exogenous but predetermined. An example in the context of count-panel-data models is the specification of Hall et al. (1984) in which patents depend on present and past R&D expenditures. It is well known that new patents will generate the need for future R&D expenditures (development).

Let us consider the conditional expectation specification

$$E(y_{it}|x_i^t, \alpha_i) = \exp(\alpha_i + \beta_0' x_{it}), \quad (24)$$

where $x_i^t \equiv (x_{it}, x_{i,t-1}, \ldots, x_{i0})$. Condition (24) is weaker than (7), given that the conditioning refers only to past values of x. In essence this means that x is predetermined and not strictly exogenous.

The nonlinear specification of the conditional expectation eliminates the possibility of using the transformation proposed in the literature for the case of linear panel data with predetermined explanatory variables. The general approach to the problem is similar, however. The basic idea is to transform (24) in a way that eliminates the unobservable effects and then use all the valid moment conditions to obtain a GMM estimator. Wooldridge (1991) presented an alternative approach to the transformation developed in the following pages.

The conditional expectation (24) could be written in generic form as

$$E(u_{it}|x_i^t) = E(d_t(y_i, x_i, \beta_0) - r_t(x_i^t, \beta_0)c_i|x_i^t) = 0, \quad (25)$$

where r_t and d_t are given functions, $r(\cdot,\cdot)$ depends only on x_i^t and β_0, and c_i is the unobservable fixed effect. Then, the transformation

$$\psi_t(y_i, x_i, \beta) \equiv d_t(y_i, x_i, \beta) - r_t(x_i^t, \beta) r_{t'}(x_i^{t'}, \beta)^{-1} \times d_{t'}(y_i, x_i, \beta) t' > t \quad (26)$$

eliminates the unobservable individual effect and has an expected value equal to 0 conditional on x_i^t. To show that we substitute $d_t(\cdot)$ by its expression as implied by Equation (25),

$$\begin{aligned}\psi_t(y_i, x_i, \beta) &= r_t(x_i^t, \beta)c_i + u_{it} \\ &\quad - r_t(x_i^t, \beta) r_{t'}(x_i^{t'}, \beta)^{-1}[r_{t'}(x_i^{t'}, \beta)c_i + u_{it'}] \\ &= u_{it} - r_t(x_i^t, \beta) r_{t'}(x_i^{t'}, \beta)^{-1} u_{it'}, \quad t' > t.\end{aligned} \quad (27)$$

It is clear from this step that the transformation does not depend on the unobservable c_i. In addition, the conditional expectation can be calculated as

$$\begin{aligned}&E[\psi_t(y_i, x_i, \beta_0)|x_i^t] \\ &= E[u_{it}|x_i^t] - E[r_t(x_i^t, \beta_0) r_{t'}(x_i^{t'}, \beta_0)^{-1} \\ &\quad \times E[u_{it'}|x_i^{t'}]|x_i^t] = 0.\end{aligned} \quad (28)$$

A particular specification of (25) that is appropriate for models with nonnegative endogenous variables and predetermined explanatory variables can be obtained as follows. First, Equation (25) can be particularized for the case we are interested in just by seeing that

$$d_t(y_i, x_i, \beta) = y_{it}, \quad r_t(x_i^t, \beta) = \exp(\beta' x_{it}),$$

$$c_i = \exp(\alpha_i). \quad (29)$$

From (26) the transformation $\psi(\cdot)$ in this case takes the form

$$\psi_t(y_i, x_i, \beta) = y_{it} - y_{it'} \exp(\beta'(x_{it} - x_{it'})), \quad t' > t. \quad (30)$$

The conditional expectation of (30) is

$$\begin{aligned}&E[\psi_t(y_i, x_i, \beta_0)|x_i^t] \\ &= E[(u_{it} - u_{it'} \exp(\beta_0'(x_{it} - x_{it'})))|x_i^t] \\ &= -E[E[u_{it'}|x_i^{t'}] \exp(\beta_0'(x_{it} - x_{it'}))|x_i^t] = 0.\end{aligned} \quad (31)$$

Transformation (30) was proposed by Chamberlain (1992) in the context of sequential moment restrictions for models with multiplicative fixed effects.

Set $t' = t + 1$. Then the transformation in Equation (30) becomes

$$\psi_t(y_i, x_i, \beta) = y_{it} - y_{it+1} \exp(\beta'(x_{it} - x_{it+1})) \quad (32)$$

Define

$$\psi_i(\beta) \equiv \begin{bmatrix} \psi_{i2}(\beta) \\ \psi_{i3}(\beta) \\ \cdots \\ \psi_{iT}(\beta) \end{bmatrix}, \quad (33)$$

where $\psi_{it} \equiv \psi(y_{it}, x_{it}, \beta)$. The matrix of instruments can be written as

$$Z_i \equiv \begin{bmatrix} z_{i2} & 0 & \cdots & 0 \\ 0 & z_{i3} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & z_{iT} \end{bmatrix}, \quad (34)$$

where $z_{it} = f(x_i^t)$. Contrary to the case in which the variables are strictly exogenous, in this case there is no common set of valid instruments. Instead, the set increases with the number of periods.

Using this notation and transformation (30),

$$E[Z_i' \psi_i(\beta_0)] = 0. \quad (35)$$

Given the choice of instruments, the GMM estimator, $\hat{\beta}$, is obtained by solving the expression

$$\min_{\beta} \left[\sum_{i=1}^{N} Z_i' \psi_i(\beta)\right]' \hat{W}_k^{-1} \left[\sum_{i=1}^{N} Z_i' \psi_i(\beta)\right], \quad (36)$$

where $\hat{W}_k$ is calculated as

$$\hat{W}_k = \frac{1}{N} \sum_{i=1}^{N} Z_i' \psi_i(\hat{\beta}_k) \psi_i(\hat{\beta}_k)' Z_i \quad (37)$$

and $\hat{\beta}_k$ is the kth round estimator.

The asymptotic variance–covariance matrix of $\hat{\beta}$ is estimated by

$$\hat{V}(\hat{\beta}) = \frac{1}{N} [D(\hat{\beta})' \Omega(\hat{\beta})^{-1} D(\hat{\beta})]^{-1}, \quad (38)$$

where

$$D(\hat{\beta}) = \frac{1}{N} \sum_{i=1}^{N} Z_i' \nabla \psi_i(\hat{\beta}) \quad (39)$$

and

$$\Omega(\hat{\beta}) = W(\hat{\beta}). \quad (40)$$

4. APPLICATIONS

Two empirical applications will be used to illustrate situations in which the procedures discussed in Section 3 are suitable. One analyzes the determinants of the number of licensing agreements signed yearly by Japanese firms. The other is the relationship between patents and R&D expenditures in the United States at the firm level.

4.1 Application I: Technology Transfer to Japanese Firms

The motivation for this application is Montalvo and Yafeh's (1994) analysis of technology transfer to Japan. I construct a model to explain the number of licensing agreements signed by each Japanese firm in any given year. The model postulates a positive relationship between licensing agreements, sales, and technological opportunity in the sector at which a particular firm belongs, and a negative relationship between any proxy for liquidity constraints and licensing agreements.

The sample contains 461 firms during the period 1977–1981. All of these firms were listed on the Tokyo Stock Exchange (TSE). The financial variables have been obtained using the information contained in firm reports to the TSE. The manufacturing firms are in the chemicals, metals, machinery, electronics, and transportation equipment sectors. None of the firms in the sample were subject to major ownership changes (merger or takeover) or fiscal-year changes during the period of observation. No foreign-owned firms or subsidiaries are included.

Two data sources are used. The dependent variable, the number of licensing agreements signed in each fiscal year (*BIC*), is taken from the *Annual Report on Technology Imports,* published by the Science and Technology Agency (Gijutsu Donyu Nenji Hokoku). The explanatory financial variables are taken from firm reports to the TSE. Firm size is measured by sales (*SALES*). The unit of measure for this variable is hundreds of billions of yen. The results of the estimation do not change if, instead of an absolute measure of size like sales, we use a relative measure like market share at the level of three-digit industries. *S*2 defines the sales squared and represents a possible economy or diseconomy of scale in the adoption of technologies. Cash flow (*CF*) is measured in billions of yen and does not include royalty payments for licensed technology. We use this variable as a proxy for liquidity constraints [see Montalvo and Yafeh (1994) for a discussion on other possible proxies for liquidity constraints]. The variable *DDEBT* measures the change in debt for each firm.

In addition to the preceding variables, one needs to control for sector-specific technological opportunities. This variable reflects the availability of innovations for adoption in a particular industry. For instance, in the computer industry there are many more opportunities for adoption of new technologies than in the steel sector. As a practical matter, technological opportunity is defined as the number of agreements signed by other firms within the same three-digit industry (*NOLI*). Of course, to avoid simultaneity problems, the number of licensing agreements of the firm itself has not been included in the variable that represents technological opportunities. A similar approach to measure this variable in which the unweighted sum of R&D spending within the same industry was treated as a measure of technological environment was used by Bernstein and Nodiri (1989). An alternative approach was suggested by Jaffe (1986), who used a "Euclidean distance" measure to identify the distance between research activities of firms.

Montalvo and Yafeh (1994), using econometric techniques for pooled count data (Poisson regression, the negative binomial model, and the truncated Poisson model), found the estimates to be consistent with their economic model: Sales, cash flow, and "technological opportunity," as defined previously, have a positive effect on the number of licensing agreements. The variable "sales squared" has a negative sign, reflecting some diseconomies of scale in the adoption of new technologies.

A natural extension of the pooled data procedures used by Montalvo and Yafeh (1994) consists of using a count-panel-data method, given the fact that the sample contains many firms over several periods of time. Table 1 shows the results of the CMLE presented in Section 1. The numbers in parentheses are the asymptotic standard deviations calculated using the estimated information matrix

$$\hat{VCV}(\hat{\beta}) = \left[\sum_{i=1}^{N} n_i \nabla_\beta \hat{p}'_{it} \begin{bmatrix} \hat{p}_{i1} & 0 & \cdots & 0 \\ 0 & \hat{p}_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{p}_{iT} \end{bmatrix} \nabla_\beta \hat{p}_{it}\right]^{-1}. \tag{41}$$

The figures between squared brackets are robust estimates of the standard deviation of the coefficients calculated as

$$\hat{\Lambda} = \frac{1}{N}[\hat{\Delta}'\hat{\Phi}^{-1}\hat{\Delta}]^{-1}, \tag{42}$$

where

$$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N} \nabla_\beta \psi_i(\hat{\beta}). \tag{43}$$

Table 1. CML Estimator for the Panel

Variable	1	2	3
SALES	.102	.121	.094
	(.026)	(.023)	(.020)
	[.078]	[.087]	[.080]
S2	−.003	−.003	−.003
	(.0008)	(.0006)	(.0006)
	[.002]	[.002]	[.002]
NOLI	.008	.008	.009
	(.001)	(.001)	(.001)
	[.006]	[.006]	[.006]
CF	—	−.003	—
	(.001)		
	[.007]		
DDEBT	—	—	.002
		(.001)	
		[.006]	
MLL	−3.8818	−3.8813	−3.8815

$$\hat{\Phi} = \frac{1}{N} \sum_{i=1}^{N} \psi_i(\hat{\beta})\psi_i(\hat{\beta})', \qquad (44)$$

and $\psi_i(\beta)$ is defined as in Equation (8).

The estimates in Table 1 are essentially consistent with the model outline by Montalvo and Yafeh (1994). The only puzzle refers to the effect of cash flow on the number of licenses. Montalvo and Yafeh (1994) found that any variable that controls for possible liquidity constraints, reflecting a lower cost of capital, has a positive effect on the number of licenses. The sign of the variable *CF* in Table 1, however, contradicts this result: It is negative and significantly different from 0, at least according to the estimates of the standard deviation derived from the information matrix.

To explain why these results are different from the ones obtained using pooled-count-data procedures, we need to analyze the conditions that justify the use of these alternative estimates. The consistency of the pooled-data estimator depends on two conditions, the distributional assumptions and the fact that $E(\alpha_i|x_t) = E(\alpha_i)$. The fixed-effect estimator is not necessarily more robust than the pooled estimator unless the strict-exogeneity assumption holds. Therefore, any significant difference between the estimates using the pooled data and the fixed panel data can be regarded as a violation of the strict-exogeneity assumption and/or the distributional assumptions underlying the individual effects.

In fact, in the context of technology transfer, it is difficult to maintain the strict-exogeneity assumption for *SALES* when the variable to be explained is the number of licensing agreements. The reason is that when a firm buys a license it is because it expects to increase sales in the future using this new technology. Therefore, the variable *SALES* is not strictly exogenous but predetermined: The history of licenses obtained in the past affects sales in the present and the future.

The GMM estimator proposed in Section 3 is appropriate to deal with the issue of predetermined variables in the context of count-panel data. In addition, that estimator relies on a conditional-expectation condition and, therefore, there is no need to specify any strong distributional assumption. Table 2 shows the results of using the GMM estimator derived from transformation (32). The initial values were taken from the CMLE parameter estimates. The weight matrix was iterated until the probability of the minimum chi-squared statistic between two rounds was reasonably small. The initial weight matrix was taken to be the identity matrix.

In column IV1 the GMM estimator uses as Z matrix the past values of *SALES*, $z_i^t = [x_{it}, x_{it-1}, \ldots, x_{i0}]$ and the values of *S2 NOLI*, and *CF*. Column IV2 is constructed assuming that all variables should be taken as predetermined and, therefore, the past values of all the variables are included. Column IV3 presents the results when the matrix Z includes only the first valid instrument for all the variables. In this case the number of instruments does not increase with $t\,(z_i^t = [x_{it}])$. Finally, IV4 includes the results when past, present, and future values of all the variables are included in Z. The probability associated with the chi-squared tests for the overidentifying restrictions is shown in the last row of the table. This statistic can detect misspecified functional forms and instruments that are not appropriately orthogonal to the residuals.

The results in Table 2 are essentially compatible with Montalvo and Yafeh (1994). The variable *SALES* has a positive effect on the number of licensing agreements signed by a firm, and the value of this parameter, in general, is closer to the pooled-count-data models than the CMLE estimates. *S2* has a negative effect, and technological opportunity has a positive effect. With respect to cash flow, *CF*, it is positive and significantly different from 0 only in IV2. Therefore, there is no clear evidence of a significant effect of *CF* over licensing agreements. In this case, however, the coefficient of *CF* is not negative and significantly different from 0 as it was in the CML estimation. The reason could be that cash flow is not a good proxy for actual liquidity constraints. Montalvo and Yafeh (1994) used the absence of bank in the financial group as a better proxy for liquidity constraints. Because no firm changes group during the period, this variable cannot be used in the panel-data context.

The fact that the specification IV4 is clearly rejected points out that the strict exogeneity of the explanatory variables does not seem to be a reasonable assumption for this application. A formal test of the null of strict exogeneity versus the alternative of predetermined instruments can be constructed following Schmidt, Ahn, and Wyhowsky (1992). The difference between the GMM estimator in column IV4 and the one in column IV2 is the fact that the instruments in IV2 are a subset of the instruments in IV4. Therefore, the Hausman test that compares both estimators will be asymptotically chi-squared with k df, where

Table 2. GMM Estimation

Variables	*IV1*	*IV1*	*IV2*	*IV2*	*IV3*	*IV3*	*IV4*	*IV4*
SALES	.325	.138	.377	.269	.341	.137	.209	.231
	(.117)	(.131)	(.118)	(.123)	(.125)	(.136)	(.079)	(.080)
S2	−.008	−.008	−.008	−.010	−.008	−.007	−.006	−.007
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
NOLI	.012	.012	.009	.006	.011	.012	.017	.011
	(.004)	(.004)	(.003)	(.002)	(.004)	(.004)	(.003)	(.002)
CF	—	.015	—	.012	—	.015	—	.003
		(.013)		(.005)		(.014)		(.005)
Prob.	.09	.21	.21	.15	.14	.32	.58	.64

NOTE: Standard errors are shown in parentheses below parameter estimates.

k is the dimension of the vector β. The only condition is that the number of additional instruments under the null, $T(T-1)h/2$, is larger than k, where h is the number of instruments. For both cases, with and without the estimation of the parameter associated with CF, this test rejects the null of strict exogeneity [$\chi^2(4) = 9.66$ and $\chi^2(3) = 6.60$, respectively].

4.2 Application II: The Relationship Between Patents and R&D

Technical change is one of the driving forces of economic growth. Much technical change is the product of investment in research and development. One of the few measurable indicators of the success of this activity is the number and kind of patents granted to each firm. For these reasons, there has been much research on the relationship between R&D investment at the firm level and patents. A different question is to what extent patent counts are good indicators of R&D output.

Pakes and Griliches (1984) analyzed this issue using a distributed-lags model of patents on past R&D. In their panel-data model, the contemporaneous R&D and the fifth lag of R&D have significant effect on patents, but only for the research-intense firms. In the case of less research-intense firms, only contemporaneous R&D has a significant effect on patents.

Hall et al. (1984) and Hall et al. (1986) considered the same question, but they used econometric techniques appropriate for count data to deal with the discrete and nonnegative nature of the number of patents, the dependent variable.

Hall et al. (1986) worked with two samples that are larger than the one used by Hall et al. (1984). The criterion selection for both of them was based on the absence of jumps in the data on sales, gross capital, and market value and on R&D being available for all the years. The first sample ($S1$) contains 642 firms with R&D data between 1972 and 1979. The second sample ($S2$) has less firms, 346, but a longer history of R&D data than the first one, covering the period 1970–1979.

Table 3. Comparison of Alternative Estimators

Case	HHG	CMLE	PMLC	GMM
$R\&D_t$	.31	.32	.30	.41
	(.04)	(.02)	(.10)	(.26)
$R\&D_{t-1}$	.02	−.09	−.10	.23
	(.05)	(.02)	(.08)	(.10)
$R\&D_{t-2}$	.03	.03	.06	−.11
	(.06)	(.02)	(.06)	(.11)
$R\&D_{t-3}$	.07	.05	−.0005	−.03
	(.06)	(.02)	(.06)	(.08)
$R\&D_{t-4}$	−.06	−.003	.06	−.04
	(.07)	(.02)	(.07)	(.08)
$R\&D_{t-5}$	−.03	−.005	.04	.09
	(.05)	(.02)	(.07)	(.14)
$\sum R\&D_{t-j}$	.43	.31	.36	.56
		(.12)	(.23)	
Trend	−.03	−.05	−.06	−.09
	(.003)	(.01)	(.01)	(.02)

NOTE: Standard errors are shown in parentheses below parameter estimates.

Table 3 shows the results of different estimation procedures. Following most of this literature, the specification of the relationship between patents and R&D is taken to be a distributed-lags model in all cases. The first column, HHG, replicates the results of the CMLE for the sample of 128 firms used by Hall et al. (1984).

To make the results comparable to ours, column CMLE presents the conditional maximum likelihood estimator based on the second sample ($S2$). Based also on this sample, column PMLC presents the results for the pseudo(quasi) maximum likelihood estimator with correlated effects (Gourieroux et al. 1984a; Hall et al. 1986).

The consistency of the previously described panel-data estimators relies on the strict exogeneity of the expenditure in R&D with respect to patents. Current and past R&D expenditures produce patents. Once a patent is granted, however, or even applied for, there is an implicit forthcoming need for additional R&D expenditures that transform patents into benefits or finance research of the new technological opportunities opened by the new patented procedures. Therefore, R&D investment should be considered a predetermined variable instead of a strictly exogenous one. For this reason, column GMM shows the result obtained using the GMM estimator presented in Section 3. The conditional expectation is specified as

$$E(\text{pat}_{it}|\alpha_i, x_i^t) = \exp\left(\alpha_i + \sum_{j=0}^{5} \beta_j \log(\text{R\&D})_{i,t-j}\right), \quad (45)$$

where pat is the number of patents and R&D is the size of R&D expenditures.

As we see in column HHG, for the fixed-effects CMLE only contemporaneous R&D has a significant effect on patents. The increase in the sample size from 128 firms to 346 makes the first lag significantly different from 0 but negative. It also reduces the total effect of past R&D on patents with respect to the pooled Poisson regression and the CMLE applied to the small sample.

The use of the PMLE with correlated effect recovers the results of the small-sample CMLE estimates: Only contemporaneous R&D has effect on patents. In addition, the effect of the sum of coefficients on past R&D is reduced.

Finally, column GMM tells a somehow different story about the influence of past R&D on patents. Contemporaneous R&D is barely significantly different from 0, but the first lag of R&D is now positive. In addition, the total effect of R&D on patents is larger than using the alternative fixed-effect panel-data estimators, although it is not as large as the one derived from the pooled Poisson regression.

5. CONCLUDING REMARKS

The "traditional" approach to estimate count-panel-data models is the CMLE. Under the Poisson assumption, the conditional distribution of the dependent variable with respect to the total number of events over the whole period of analysis is a multinomial distribution. In the panel-data context, this approach is particularly useful because conditional distribution does not depend on the unobservable effects.

The CMLE is strongly required to be consistent, however. In particular, the strict exogeneity of the explanatory variables may be a strong assumption in the context of many economic applications. We develop a GMM estimator, based on a transformation proposed by Chamberlain (1992) to estimate a count-panel-data model without imposing strict exogeneity or any particular distributional assumption. This method is applied to the explanation of technology transfer to Japanese firms and the relationship between patents and R&D expenditures at the firm level. In both cases the explanatory variables seem to be predetermined instead of strictly exogenous.

ACKNOWLEDGMENTS

I thank G. Chamberlain, B. Hall, Z. Griliches, G. Imbens, D. Jorgenson, Y. Yafeh, seminar participants at Harvard, Rochester, Boston University, and the National Bureau of Economic Research Summer Institute, two anonymous referees, and, especially, the associate editor. Financial support by Grant DGCYT PB93/0833 from the Spanish Ministry of Education is gratefully acknowledged. I also thank B. Hall and Y. Yafeh, who provided me with the data used in the applications. This is a revised version of the paper "Multiplicative Panel Data Models: An Application to Technology Transfer."

[Received April 1993. Revised September 1995.]

REFERENCES

Andersen, E. B. (1970), "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, Ser. B, 32, 283–301.

Anderson, T. W., and Hsiao, C. (1981), "Estimation of Dynamic Models With Error Components," *Journal of the American Statistical Association*, 76, 598–606.

Arellano, M., and Bond, S. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and Applications to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Bernstein, J., and Nadiri, I. (1989), "Research and Development and Intraindustry Spillovers: An Empirical Application of Dynamic Duality," *Review of Economic Studies*, 56, 2, 52–78.

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

Chamberlain, G. (1992), Comment on "On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," by M. P. Keane and D. E. Runkle, *Journal of Business & Economic Statistics*, 10, 20–26.

Efron, B. (1992), "Poisson Overdispersion Estimates Based on the Method of Asymmetric Maximum Likelihood," *Journal of the American Statistical Association*, 87, 98–107.

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

——— (1984b), "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701–720.

Hall, B. H., Griliches, Z., and Hausman, J. (1986), "Patents and R&D: Is There a Lag?" *International Economic Review*, 27, 265–283.

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

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

Hayashy, F., and Sims, C. (1983), "Nearly Efficient Estimation of Time Series Models With Predetermined but not Exogenous Instruments," *Econometrica*, 51, 783–792.

Holtz-Eakin, D., Newey, W., and Rosen, H. (1988), "Estimating Vector Autoregressions With Panel Data," *Econometrica*, 56, 1071–1395.

Jaffe, A. (1986), "Technological Opportunity and Spillovers of R&D: Evidence From Firms Patents and Market Value," *American Economic Review*, 76.

Keane, M. P., and Runkle, D. E. (1992), "On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," *Journal of Business & Economic Statistics*, 10, 1–9.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

Montalvo, J. G., and Yafeh, Y. (1994), "A Microeconometric Analysis of Technology Transfer: The Case of Licensing Agreements of Japanese Firms," *International Journal of Industrial Organization*, 12, 227–244.

Mullahy, J. (1986), "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics*, 33, 341–365.

Pakes, A., and Griliches, Z. (1984), "Patents and R&D at the Firm Level: A First Look," in *R&D, Patents and Productivity*, ed. Z. Griliches, Chicago: University of Chicago Press, pp. 55–72.

Schmidt, P., Ahn, S. C., and Wyhowski, D. (1992), Comment on "On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," by M. P. Keane and D. E. Runkle, *Journal of Business & Economic Statistics*, 10, 10–14.

Wooldridge, J. M. (1990), "Distribution Free Estimation of Some Nonlinear Panel Data Models," unpublished manuscript, Massachusetts Institute of Technology.

——— (1991), "Multiplicative Panel Data Models Without the Strict Exogeneity Assumption," unpublished manuscript, Massachusetts Institute of Technology,