NBER TECHNICAL PAPER SERIES

ECONOMETRIC MODELS FOR COUNT DATA WITH AN APPLICATION TO THE PATENTS-R&D RELATIONSHIP

Jerry A. Hausman

Bronwyn Hall

Zvi Griliches

Technical Paper No. 17

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

August 1981

This paper was presented at the NBER 1981 Summer Institute in the Productivity program. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #17
August 1981

Econometric Models for Count Data with an Application to the Patents-R&D Relationship

ABSTRACT

This paper focuses on developing and adapting statistical models of counts (non-negative integers) in the context of panel data and using them to analyze the relationship between patents and R&D expenditures. The model used is an application and generalization of the Poisson distribution to allow for independent variables; persistent individual (fixed or random) effects, and "noise" or randomness in the Poisson probability function.

We apply our models to a data set previously analyzed by Pakes and Griliches using observations on 128 firms for seven years, 1968-74. Our statistical results indicate clearly that to rationalize the data, we need <u>both</u> a disturbance in the conditional within dimension and a different one, with a different variance, in the marginal (between) dimension. Adding firm specific variables, log book value and a scientific industry dummy, removes most of the positive correlation between the individual firm propensity to patent and its R&D intensity. The other new finding is that there is an interactive negative trend in the patents - R&D relationship, that is, firms are getting less patents from their more recent R&D investments, implying a decline in the "effectiveness" or productivity of R&D.

Jerry A. Hausman
Department of Economics
Massachusetts Institute of Technology
Cambridge, Mass. 02139

(617) 253-3644

Bronwyn Hall
National Bureau of Economic Research
204 Junipero Serra Blvd.
Stanford, CA 94305

(415) 326-7160

Prof. Zvi Griliches
National Bureau of Economic Research
1050 Massachusetts Ave.
Cambridge, Mass. 02138

(617) 868-3921
(617) 495-2181

July 1981

Econometric Models for Count Data with an Application to the Patents-R&D Relationship

Jerry Hausman, Bronwyn Hall, Zvi Griliches*

Introduction

This paper arose out of the analysis of a specific substantive problem: the relationship between the research and development (R&D) expenditures of firms and the number of patents applied for and received by them. There are two salient aspects of the data we wish to analyze. 1) Our dependent variable is a count of the total number of patents applied for by a particular firm in a given year. It varies from zero to several or even many, for some firms. 2) We have repeated observations for the same firms. That is, our data form a combined time-series cross-section panel. In this paper, we focus, therefore, on developing and adapting statistical models of counts (non-negative integers) in the context of panel data and using them to analyze the relationship between patents and R&D expenditures. This is not, however, the only possible application for the methods discussed in this paper. A variety of other economic data come in the form of repeated counts of some individual actions or events. The number of annual visits to dentists, the number of records purchased per month, the number of cars owned, or the number of jobs held during a year, all have non-negligible

*MIT, NBER, Palo Alto, and Harvard University, respectively.
We are indebted to NSF Grants SES79-24108, SOC78-04279, and PRA79-13740 for financial support of this work.

probabilities of zero and are non-negative integers.

The statistical models we develop are applications and generalizations of the Poisson distribution. After rewriting the Poisson distribution as a function of a number of independent variables we have to deal with two additional issues: (1) Given the panel nature of our data, how can we allow for separate persistent individual (fixed or random) effects? and (2) How does one introduce the equivalent of disturbances-in-the-equation into the analysis of Poisson and other discrete probability functions?

The first problem is solved by conditioning on the total sum of outcomes over the observed years, while the second problem is solved by introducing an additional source of randomness, allowing the Poisson parameter to be itself randomly distributed, and compounding the two distributions. The relevant likelihood functions and the associated computational methods are described in the body of the paper.

The substantive application continues the work of Pakes and Griliches (1980a and b). In that work patent data for 8 years (1968-1975) and 121 U.S. companies were analyzed as function of their current and lagged R&D expenditures. A log-log functional form was used and the "zero value" problem was "solved" by (a) choosing companies so as to minimize this problem (only 8 percent of the observations were zero in any one year) and (b) setting zeroes equal to one and adding a dummy variable to allow the equation to choose implicitly another value between zero and one. The questions of interest were (a) the strength (fit) of the relationship between patents and R&D, (b) the elasticity of patents with respect to R&D expenditures, (c) the shape of the distributed lag of R&D effects, and (d) the presence and sign of a trend in this relationship. The major findings were: A high fit ($R^2 \cong .9$) cross-sectionally and a lower ($R^2 \cong .3$) though still statistically significant fit in the "within" time series dimension of the data. The estimated elasticity was around 1.0 in the cross-sectional dimension, dropping to about .5 in the within, shorter-run time dimension. The shape of the distributed lag was not well defined, with some indication of lag-truncation bias (the possible influence of pre-sample unmeasured R&D expenditures) which could not, however, be well distinguished from a fixed firm effect.* A negative time trend was found in most of the examined data subsets.

In this paper we wish to reexamine the earlier findings using a more appropriate model for such data, a model that reflects explicitly its integer nature. We do not expect the results to change much since the "zero" problem is relatively minor in this sample (8 percent). We are interested, however, in developing this methodology because the sample is being expanded to encompass many more smaller firms with a concomitant increase in the importance of such issues. We use a sample of 128 firms for the 7 years 1968-1974. The patent data were tabulated for us by the Office of Technology Assessment and Forecasting of the U.S. Patent Office and the R&D data were taken from the Computstat tape and other sources (see Pakes and Griliches, 1980, for more detail on sample derivation and construction), and deflated by an approximate R&D cost deflator.

---

*It is difficult to distinguish in a short time series between a left-out pre-sample cumulated R&D value whose effect is dying out slowly and a "permanent fixed" individual firm effect. See Griliches and Pakes, 1980, for further discussion of these issues.

The rest of the paper is organized as follows: Section 1 presents the simple Poisson regression model and applies it to our data. Section 2 develops a generalization which allows each firm to have its own average propensity to patent by conditioning separately the count distribution of each firm on the sum of its patents for the whole period. Section 3 allows for the "over-dispersion" in the data by letting each firm's Poisson parameter have random distribution of its own, leading to the estimation of a negative-binomial model for these data. Section 4 explores in our non-linear context the parallels to the "within" - "between" dichotomy in linear models. Section 5 summarizes the major methodological and substantive results and discusses some possible future lines of work.

1. The Poisson Model and Application

The Poisson distribution is often a reasonable description for events which occur both "randomly and independently" in time.[1] It seems a natural first assumption for many counting problems in econometrics. Let us denote the Poisson parameter as λ, and consider specifications of the form $\log \lambda = X\beta$ where X is a vector of regressors which describe the characteristics of an observation unit in a given time period. Denote n_{it} as the observed event count for unit i during the time period t. The advantages of the Poisson specification are: (1) in many ways it is analogous to the familiar econometric regression specification. In particular, $E(n_{it}|X_{it}) = \lambda_{it}$. Furthermore, estimation of unknown parameters is straightforward and is done either by an iterative weighted least squares technique or by a maximum likelihood algorithm. The log likelihood function is globally concave so that maximization routines converge rapidly. (2) The "zero problem," $n_{it} = 0$, is a natural outcome of the Poisson specification. In contrast to the usual logarithmic regression specification we need not truncate an arbitrary continuous distribution. Likewise, the integer property of the outcomes n_{it} is handled directly. For large n_{it} a continuous approximation often suffices. But for small n_{it}, a specification which models the counting properties of the data (both large and small) seems in order. (3) The Poisson specification allows for convenient time aggregation so long as its basic assumption of time independence holds true. Thus, if the counting process is Poisson over time t=1,T with parameter λ_{it}, then the aggregate data over period zero to T are also Poisson with parameter $\lambda_{it} = \sum_{t=1}^{T} \lambda_{it}$.

This property permits the convenient generalization of the Poisson model to be developed below. The time independence property is also a potential weakness of our specification given the often noted serial correlation of residuals in econometric specifications. We will attempt to distinguish carefully between true time independence versus apparent dependence due to unobserved heterogeneity of the individual units.

Our basic Poisson probability specification is

$$(1.1) \qquad pr(n_{it}) = f(n_{it}) = \frac{e^{-\lambda_{it}} \lambda_{it}^{n_{it}}}{n_{it}!}$$

In our application, i indexes firms and t indexes years and we specify $\log \lambda_{it} = X_{it}\beta$. Note that λ_{it} is a deterministic function of X_{it}, and the randomness in the model arises from the Poisson specification for the n_{it}. The moment generating function of the Poisson distribution is $m(t) = e^{-\lambda}e^{\lambda e^{t}}$ so that the first two moments are $E(n_{it}) = \lambda_{it}$ and $V(n_{it}) = \lambda_{it}$. The regression property of this specification arises from $E(n_{it}) = \lambda_{it}$, but it is not uncommon to find that the variance of n_{it} is larger than the mean empirically, implying "overdispersion" in the data. After an initial exploration of the Poisson model, we shall consider the possibility of such overdispersion.

The log likelihood of a sample of N firms over T time periods for the Poisson specification is

$$L(\beta) = \sum_{i=1}^{N} \sum_{t=1}^{T} [-\lambda_{it} + n_{it} \log\lambda_{it} - n_{it}!]$$

(1.2)

$$= \sum_{i=1}^{N} \sum_{t=1}^{T} [C_1 - e^{X_{it}\beta} + n_{it}X_{it}\beta]$$

where C_1 is a constant. The gradient and Hessian take the forms

$$\frac{\partial L}{\partial \beta} = \Sigma\ \Sigma\ [X_{it}(n_{it} - e^{X_{it}\beta})]$$

(1.3)

$$\frac{\partial^2 L}{\partial\beta\partial\beta'} = \Sigma\ \Sigma\ [-(X'_{it}X_{it})e^{X_{it}\beta}].$$

The first order conditions indicate that an iterative nonlinear weighted least squares program with $n_{it} - \lambda_{it}$ as the "residual" could be used to estimate β by maximum likelihood (ML). The Hessian demonstrates that the likelihood function is globally concave so long as X is of full column rank and $e^{X_{it}\beta}$ does not go to zero for all X_{it}. With a globally concave likelihood function, a wide choice of ML algorithms can be used. In our applications convergence to the global maximum was always rapid. The variance matrix of the asymptotic distribution $V(\beta)$ is calculated from the Hessian matrix evaluated at β.

We fit our initial Poisson specification to a model with current R&D and five lagged values of R&D, and a time trend. The results are found in Table 1. We also present the corresponding estimates of a least squares regression of $\log(n_{it}) = X_{it}\beta + \varepsilon_{it}$ where $\log(n_{it})$ is set to zero and a dummy variable used when $n_{it} = 0$. The results of the Poisson model are broadly similar to

TABLE 1

Estimates for the Patents Model - Totals[1]

	Variable Means	Poisson	OLS[2]	Poisson	OLS	Poisson
log R_o	2.17 (1.64)	.36 (.03)	.21 (.12)	.87 (.004)	.81 (.02)	.65 (.01)
log R_{-1}	2.15 (1.62)	.13 (.05)	.12 (.17)			
log R_{-2}	2.12 (1.61)	.09 (.05)	-.05 (.18)			
log R_{-3}	2.08 (1.62)	-.13 (.06)	-.08 (.18)			
log R_{-4}	2.03 (1.63)	.11 (.07)	.01 (.19)			
log R_{-5}	1.96 (1.64)	.53 (.05)	.64 (.15)			
time		-.06 (.002)	-.06 (.02)	-.04 (.002)	-.009 (.015)	.04 (.01)
time·log R_o						-.02 (.002)
dummy[3] (scientific sector)	0.52					.37 (.01)
log book value[4]	5.36 (2.00)					.24 (.004)
intercept		1.68 (.02)	1.58 (.08)	1.55 (.018)	1.32 (.08)	.39 (.04)
dummy (n_{it} = 0)	0.08		-1.37 (.12)		-1.41 (.12)	
sum of log R coefficients		.88	.84	.87	.81	.57[5]
standard error			.869		.914	
log likelihood		-10,621.1		-11,198.4		-9,077.5

[1] The sample is 128 firms, annual data from 1978 to 1974. In column 1, standard deviations are in parentheses.

[2] For the OLS estimates, the dependent variable is log of patents, with a dummy for observations with zero patents.

[3] The scientific sector dummy is for firms in the drug, computer, scientific instrument, chemical, and electronic equipment industries.

[4] The log book value variable is the natural logarithm of the inflation adjusted book value of the firms in 1971.

[5] Evaluated at mean time = 4.

Table la

Correlation of the Standardized Residuals for the Poisson Model

	68	69	70	71	72	73	74
68	1.0						
69	.88	1.0					
70	.80	.89	1.0				
71	.74	.82	.87	1.0			
72	.69	.77	.79	.89	1.0		
73	.64	.77	.77	.82	.86	1.0	
74	.55	.62	.62	.67	.75	.82	1.0

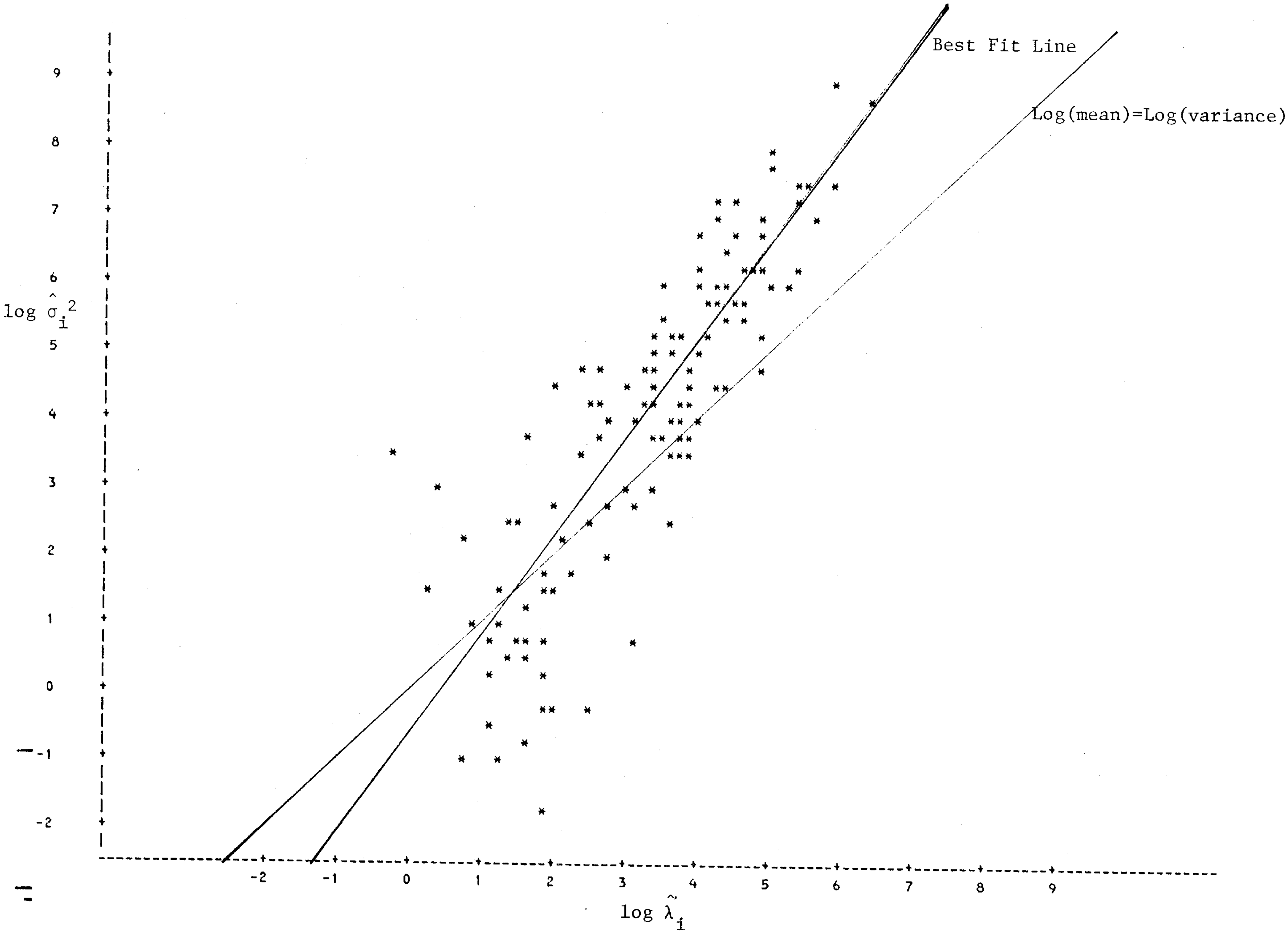
OLS although note that the estimated standard errors of the Poisson estimates are approximately three times smaller. The coefficient of current R&D is higher but the sum of the lag coefficients are quite similar. We note an exogeneous decrease in patents of 6% per year. Lastly, we have the somewhat disturbing pattern of a U-shaped distributed lag which may well indicate a substantial truncation effect. This problem will be eliminated shortly when we specify firm specific effects.

We now consider alternative specifications of the basic Poisson model. In column 4 of Table I we delete all lagged R&D. It turns out that when firm specific effects are added that the lagged effects become quite small. Note that the coefficient of current R&D is very close to the sum of the coefficients in our initial specification. The exogenous time effect has now decreased in magnitude to 4% per year. In column 5 we find very similar though slightly lower results from the OLS regression. Lastly, in column 6 we add a dummy variable for the scientific sector which includes firms in the drug, computer, scientific instruments, chemical, and electronic equipment industries. We also add a variable for the inflation adjusted book value of the firm in 1971. Both variables have strong positive effects on the expected number of patents. In addition, we interact R&D with time to attempt to sort out a pure exogenous effect of time from a decrease in the effectiveness of R&D over time. The estimates indicate that the effect of R&D seems to be decreasing since the estimated coefficient is -.02 while the time coefficient has now switched sign to +.04. Both effects are precisely estimated and they tend to persist when we move to more elaborately specified models.

To evaluate the adequacy of the Poisson specification we now turn to an investigation of the residuals. Starting with the Poisson residual $\hat{u}_{it} = n_{it} - \hat{\lambda}_{it}$ we define the standardized residual as $\hat{u}_{it}$ divided by its estimated standard deviation: $\hat{\varepsilon}_{it} = (n_{it} - \hat{\lambda}_{it})/\sqrt{\lambda_{it}}$. We use these residuals to test our model specification in three related ways. First the independence assumption can be tested by forming the 7 x 7 covariance matrix $\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (\hat{\varepsilon}_i \hat{\varepsilon}_i')$ where ε_i is a vector of residuals for firm i. One potential problem arises here. Since a common $\hat{\beta}$ is used to form $\hat{u}$, under the null hypothesis of zero covariance of the true u_{it}'s, induced covariance of order (1/NT) exists among the $\hat{u}_{it}$'s. But since NT=896 in our sample , this problem and the associated Cox-Snell (1969) corrections are quite small . The estimated correlation matrix for the specification of column 6 of Table 1 is found in Table 1a. Note that significant correlation exists which casts serious doubts on the adequacy of our Poisson specification.[2] Next we consider the variance property. Given the Poisson specification the variance of the ε_{it}'s should be unity. In Figure 1 we show a log-log plot of $\hat{\sigma}_t^2 = \frac{1}{T-1} \Sigma(\varepsilon_{it} - \bar{\varepsilon}_i)^2$ for $\bar{\varepsilon}_i = \frac{1}{T} \sum_t \hat{\varepsilon}_{it}$ against $\bar{\lambda}_i = \frac{1}{T} \sum_t \hat{\lambda}_{it}$. We do not find the expected one-to-one relationship at all. The variance increases considerably more rapidly than does the mean. A simple regression of log $\hat{\sigma}_t^2$ on log $\bar{\lambda}_i$ takes the form, $\log \hat{\sigma}_{\varepsilon_i}^2 = -.68 + 1.42 \log \bar{\lambda}_i$. Thus, we need also turn our attention to this failure of our initial specification. Lastly, we consider the omnibus Pearson χ^2 type test for

Figure 1

WITHIN FIRM RESIDUAL VARIANCE VERSUS AVERAGE LAMBDA



goodness of fit. Because we have estimated the coefficients by maximum likelihood, the Chernoff-Lehmann (1954) problem of appropriate degrees of freedom arises. However $W = \Sigma\Sigma\varepsilon_{it}^{2} = 18{,}900$ for column 2 of Table 1 which far exceeds normal significance levels for even 896 degrees of freedom. Likewise W = 15,655 for our last specification in column 6. The Pearson test confirms our other results that a more general form than the simple Poisson specification is required for an adequate representation of the patent data.

2. Firm Specific Effects

Investigation of the standardized residuals from the Poisson estimation clearly indicates the presence of serial correlation. Such a finding is not uncommon in panel data of the type we are using. If unobserved firm specific effects exist, the residuals for a given firm might all be of the same sign indicating the way in which the firm deviates from the "average firm." Also, the finding of a U-shaped distributed lag indicates the possible presence of firm specific truncation effects. We know from the analysis of linear panel data models that there are two methods which can be used for this type of problem. We explore first the random effects specification. In the regression model this implies an equicorrelated covariance matrix and is sometimes sufficient to explain the apparent serial correlation. In our Poisson specification the random effect has somewhat similar implications. We specify $\tilde{\lambda}_{it} = \lambda_{it}\tilde{\alpha}_i$ where $\tilde{\alpha}_i$ is a random firm specific effect. The Poisson parameter $\tilde{\lambda}_{it}$ is now also a random variable rather than a deterministic function of X_{it}. Correlation of $\tilde{\lambda}_{it}$ and $\tilde{\lambda}_{it'}$, $(t \neq t')$ arises from the $\tilde{\alpha}_i$ while $\tilde{\lambda}_{it}$ and $\tilde{\lambda}_{jt}$ are uncorrelated by the assumption of independent $\tilde{\alpha}_i$.

The other approach to firm specific effects is to condition on the $\tilde{\alpha}_i$ and apply conditional maximum likelihood techniques of Andersen (1970, 1972). We then have a fixed effects specification. While asymptotic efficiency is sacrificed by the conditioning, no distribution need be specified for the $\tilde{\alpha}_i$. Perhaps more important while we might specify the $\tilde{\alpha}_i$ to be random, conditional on the X_{it} they may no longer be randomly distributed or exchangeable, using diFinetti's approach. For example, firms which are better at producing patents for unobserved reasons may invest more in R&D because they obtain a higher return to the expenditure. The random effects specification is then no longer valid.[1] We use Hausman's (1978) test to decide whether there exists a significant non-random correlation between the X_{it} and the $\tilde{\alpha}_i$'s.

We first consider the random effects specification. Note that an important difference exists between the random effects Poisson specification and the random effects regression specification. Here the estimated β's are inconsistent if the random effect is omitted from the specification. In the regression specification, on the other hand, the random effects raise only problems of efficiency. Because $\tilde{\lambda}_{it}$ needs to be positive, we write it in the form $\tilde{\lambda}_{it} = \lambda_{it}\alpha_i = e^{X_{it}\beta + \mu_i}$ where μ_i is the firm specific effect. The Poisson probability specification then becomes

$$\text{pr}(n_{it}|X_{it},\mu_i) = \frac{e^{-e^{X_{it}\beta + \mu_i}}\left(e^{X_{it}\beta + \mu_i}\right)^{n_{it}}}{n_{it}!} \tag{2.1}$$

$$= \frac{e^{-\lambda_{it}e^{\mu_i}}\left(\lambda_{it}e^{\mu_i}\right)^{n_{it}}}{n_{it}!}$$

The joint density of $(n_{i1},\ldots,n_{it})$ and μ_i takes the form

$$(2.2) \quad pr(n_{i1},\ldots,n_{iT},\mu_i|X_{i1},\ldots,X_{iT}) = pr(n_{i1},\ldots,n_{iT}|X_{i1},\ldots X_{iT}\mu_i)g(\mu_i)$$

$$= \prod_t \left[\frac{\lambda_{it}^{n_{it}}}{n_{it}!}\right] e^{-e^{\mu_i \sum_t \lambda_{it}}} \left[e^{\mu_i}\right]^{\sum_t n_{it}} g(\mu_i)$$

where $g(\mu_i)$ is the probability density function of μ. In equation (2.2) we have made the important assumption that the conditional density of μ_i given X_{it} equals the unconditional density of μ_i. Thus, the μ's are assumed to be randomly distributed across firms.

Since μ_i is an unobservable random variable we now integrate it out from equation (2.2). To do so, we assume that $\alpha_i = e^{\mu_i}$ is distributed as a gamma random variables with parameters $(\gamma,\delta)^2$. We integrate by parts to find

$$(2.3) \quad pr(n_{i1},\ldots,n_{iT}|X_{i1},\ldots,X_{iT}) = {}_0\int^{\infty} \prod \left[\frac{\lambda_{it}^{n_{it}}}{n_{it}!}\right] e^{-\alpha_i \sum_t \lambda_{it}} \alpha_i^{\sum_t n_{it}} f(\alpha_i)d\alpha_i$$

$$= \prod_t \left[\frac{\lambda_{it}^{n_{it}}}{n_{it}!}\right]\left[\frac{\delta}{\Sigma\lambda_{it} + \delta}\right] \left(\Sigma\lambda_{it} + \delta\right)^{\Sigma n_{it}} \frac{\Gamma(\Sigma n_{it} + \gamma)}{\Gamma(\gamma)}$$

where $\Gamma(\cdot)$ is the gamma function, $\Gamma(z) = {}_0\int^{\infty} t^{z-1}e^{-t}dt$ for $z > 0$. For this model the expectation of n_{it} is $\gamma\lambda_{it}/\delta$ and the variance is $\gamma\lambda_{it}(\lambda_{it} + \delta)/\delta^2$. Therefore, the ratio of the variance to the mean is now $(\lambda_{it} + \delta)/\delta$ so the smaller δ is, the larger the average ratio is. The variance to mean ratio grows with λ_{it} which is what we observed previously in the residuals. The log likelihood for this random effects specification takes the form

$$(2.5) \qquad L(\beta, \gamma, \delta) = C_2 - N\log\Gamma(\gamma) + N\gamma\log\delta + \sum_i[\sum_t n_{it}(X_{it}\beta) - (\gamma + \sum_t n_{it}) \log (\sum_t e^{X_{it}\beta} + \delta) + \log \Gamma(\sum_t n_{it} + \gamma)]$$

Maximum likelihood estimation of equation (2.5) is straightforward although we can no longer prove global concavity due to the addition of the γ and δ parameters. Evaluation of the log gamma function and its derivative (the digamma function) is akin to calculation of a logarithm on a computer. Starting values are provided by the initial Poisson estimates and guesses of the gamma parameters using the moments $E\alpha = \gamma/\delta$ and $V(\alpha) = \gamma/\delta^2$.

Results of the random effects Poisson specification are given in columns 1 to 3 of Table 2. We see that the U-shaped lag structure of R&D is somewhat attenuated from that in Table 1, but there is still a significant positive coefficient on the last lag. The ratio of the variance to the mean of the firm effect implied by the model is approximately 10, except in column 3, where we have included two firm specific variables. When we include these variables, the ratio drops almost to one. These results imply that the random effects model fails to account completely for firm effects and we proceed, therefore, to specify a fixed effects model.

We emphasized in our derivation of the random effects specification of equation (2.3) the requirement that the unconditional and conditional density of μ_i given X_{it} was identical. This requirement can be dropped when a conditional maximum likelihood approach is used to develop a fixed effects specification. But we cannot simply estimate separate μ_i parameters in equation (2.1) because for T held fixed and N large we have the incidental parameter problem and maximum likelihood need not be consistent [See Neymann and Scott (1948), Andersen (1973), and Haberman (1977).] Instead, we use the conditional maximum likelihood approach of Andersen (1970, 1972) and condition on the sum of patents $\sum_t n_{it}$. Since the Poisson distribution is a member of the exponential family, a sufficient statistic exists for $T\tilde{\lambda}_i = \Sigma\tilde{\lambda}_{it}$ and it is Σn_{it}. Since $\sum_t n_{it}$ is distributed as Poisson with parameter $\sum_t \tilde{\lambda}_{it} = \alpha_i \Sigma\lambda_{it}$, conditional maximum likelihood follows in a straightforward manner. Furthermore, it is known in the literature, e.g. Rao (1952), that the distribution of n_{it} conditional on $\sum_t n_{it}$ gives a multinomial distribution

(2.6)

$$\mathrm{pr}(n_{i1},\ldots,n_{iT}|\Sigma n_{it}) = \mathrm{pr}(n_{i1},\ldots,n_{i,T-1}, \sum_{t=1}^{T} n_{it} - \sum_{t=1}^{T-1} n_{it})/\mathrm{pr}(\Sigma n_{it})$$

$$= \frac{\dfrac{e^{-\sum_t \tilde{\lambda}_{it}} \prod_t \tilde{\lambda}_{it}^{n_{it}}}{\prod_t (n_{it}!)}}{\dfrac{e^{-\sum_t \tilde{\lambda}_{it}} (\sum_t \tilde{\lambda}_{it})^{\sum_t n_{it}}}{(\sum_t n_{it})!}} = \frac{(\sum_t n_{it})!}{\prod_t (n_{it}!)} \prod_t \left(\frac{\tilde{\lambda}_{it}}{\sum_t \tilde{\lambda}_{it}}\right)^{n_{it}}$$

Set $p_{it} = \tilde{\lambda}_{it}/(\sum_t \tilde{\lambda}_{it})$ and we have the multinomial distribution since $\Sigma p_{it} = 1$.

Table 2

Estimates of the Poisson Model with Firm Effects

	Random Effects			Fixed Effects		
	1	2	3	4	5	6
log R_o	.36 (.02)	.45 (.01)	.48 (.01)	.31 (.04)	.35 (.03)	.48 (.03
log R_{-1}	.03 (.04)			.02 (.05)		
log R_{-2}	.06 (.05)			.04 (.06)		
log R_{-3}	.08 (.05)			.07 (.06)		
log R_{-4}	-.07 (.05)			-.07 (.07)		
log R_{-5}	.13 (.03)			.07 (.05)		
time	-.04 (.01)	-.03 (.005)	.037 (.003)	-.03 (.003)	-.027 (.002)	.037 (.00
time·log R_o			-.017 (.006)			-.017 (.00
dummy (scientific sector)			.18 (.16)			-
log $book_{71}$			.32 (.04)			-
γ	1.20 (.15)	.98 (.14)	1.40 (.15)			
δ	.12 (.02)	.070(.013)	.85 (.14)			
sum of log R coefficients	.59	.45	.41*	.43	.35	.41*
log likelihood	-3827.5	-3846.18	-3779.6	-3009.4	-3014.4	-2979.0
Test for correlated firm effects				15.2	13.7	.01

Random effects $\alpha_i = e^{\mu_i}$, α_i distributed independently as a gamma random variable with parameters γ and δ^2.

Fixed effects: Estimates conditional on the sum of patents over all T years.

*"Sum" evaluated at $\bar{t} = 4$.

TABLE 2a

Estimated Correlation of the BLUS-transformed Residuals[1]

	69	70	71	72	73	74
69	1.0					
70	.21	1.0				
71	-.05	.27	1.0			
72	-.22	.07	.32	1.0		
73	-.22	-.09	.08	.41	1.0	
74	-.36	-.20	-.02	.38	.59	1.0

[1] These residuals were computed using the parameter estimates in the last column of Table 2.

Furthermore, for our particular specification we have

$$p_{it} = e^{X_{it}\beta + \mu_i} / (\sum_t e^{X_{it}\beta + \mu_i}) = e^{X_{it}\beta} / \sum_t e^{X_{it}\beta}$$

which is the so-called multinomial logit specification used by McFadden (1974) in the discrete choice problem.[3] Define the share of patents for firm i in a given year by $s_{it} = n_{it}/\sum_t n_{it}$. The logit model then explains the share of total patents in each year given the firms' total number of patents in T years.

The log likelihood function takes the form

$$(2.7) \qquad L(\beta) = C_3 - \sum_{i=1}^{N} \sum_{t=1}^{T} n_{it} \log \sum_{s=1}^{T} e^{-(X_{it} - X_{is})\beta}$$

It differs from the discrete choice likelihood function because here in general all the s_{it}'s are non-zero instead of only one non-zero value for the choice which is made. We define $Z_{its} = X_{it} - X_{is}$ and calculate the gradient and Hessian

$$(2.8) \qquad \frac{\partial L}{\partial \beta} = \sum_i \sum_t \frac{n_{it}}{\sum_s e^{-Z_{its}\beta}} \sum_s e^{-Z_{its}\beta} Z_{its}$$

$$\frac{\partial^2 L}{\partial\beta\partial\beta'} = -\sum_i \sum_t \frac{n_{it}}{Z_i} \sum_s e^{-Z_{its}\beta} Z_{its} Z'_{its}$$

$$+ \sum_i \sum_t \frac{n_{it}}{Z_i^2} \left(\sum_s e^{-Z_{its}} Z'_{its}\right)\left(\sum_s e^{-Z_{its}\beta} Z_{its}\right)'$$

where $Z_i = \sum_s e^{-Z_{its}\beta}$. The Hessian takes the usual weighted cross product form. It is globally concave since by the Cauchy inequality $\sum_i \sum_t \sum_s e^{-Z_{its}\beta} Z_{its} Z'_{its} > \sum_i \sum_t \frac{1}{Z_i} \left(\sum_s e^{-Z_{its}} Z_{its}\right)\left(\sum_s e^{-Z_{its}} Z_{its}\right)'$ so long as β remains bounded. Existing logit programs can be altered to maximize equation (2.7). Our computational experience found rapid convergence because of the global concavity property.

The results for the conditional Poisson are given in columns 4 to 6 of Table 2. In the original specification with 5 lags, only the current value of R&D contributes very much to λ_{it}. Patents lagged one year as well as the U-shape of the distributed lag no longer play a role. The firm specific effect α_i now stands for the accumulated stock of knowledge from past R&D in the firm[4] and the sum of the lag coefficients is now .43 rather than the .59 value in the random effects specification.[5]

In column 5 we estimate a model which contains only current R&D since a χ^2_5 test took the value 10 which is not significant at the 5% level. We find that current R&D has a coefficient of .35 which is 20% below the sum of the coefficients in the previous specification. The time coefficient remains at -3% per year. In column 6 we redo the specification with a time and R&D

interaction. This specification corresponds to that of column 3 where the scientific sector dummy and book value variables have been absorbed into the fixed effect. The coefficient of current R&D now rises to .48 while our earlier findings about the potency of R&D are repeated. Time itself has a positive coefficient of 4% per year while the interaction with R&D has a coefficient of -.02. We find again that the effectiveness of R&D has been diminishing with time in our data.

To test whether the firm specific effects play an important role we consider two statistics. First we do an analysis of variance type test of whether all the α's are equal to the overall constant. This is a likelihood ratio (LR) test of the conditional model in column 6 of Table 2 against the corresponding model in Table 1. Under the null hypothesis the statistic is distributed as χ^2_{125}. The LR test statistic exceeds 12,000 which is a clear rejection. For a more interesting test, we compare the random effects estimates to the fixed effects estimates using Hausman's (1978) test. For columns 1 and 4, this test leads to a statistic distributed as χ^2_7 under the null hypothesis. Our statistic equals 15.2 which leads to a rejection of the random effects model. However, when we test column 3 against column 6 we accept the null hypothesis that the firm specific effects remaining after inclusion of the scientific sector dummy and the firm size variable are independent of the X_{it}'s. The value of the statistic is .01, distributed as χ^2_3 under the null.

Lastly, we consider diagnostic tests. We can no longer disregard the induced correlation in the conditional model since $\sum_t \hat{u}_{it} = 0$ which follows from the fixed effects assumption and the definition $\hat{\lambda}_{it} = \hat{s}_{it}\Sigma n_{it}$. Thus, under the null hypothesis of no serial correlation among the u_{it}, we have serial correlation of order $(\frac{1}{T})$ among the u_{it}. To form a test for serial correlation, we again disregard the possible correlation which arises from the $\hat{\beta}$'s. Since $\hat{\beta}$ is estimated on 768 degrees of freedom, the Cox-Snell (1969) corrections are quite small. On the other hand, the correlation which arises from $\sum_t \hat{u}_{it} = 0$, remains even as N, the number of firms, becomes large.

Given the linear dependence among the $\hat{u}_{it}$, we want to transform from a seven dimensional vector $\hat{u}_i$ to a 6 dimensional vector $\tilde{\varepsilon}_i$ which under the null hypothesis would not be serially correlated but would maintain the other properties of ε_i. We need an orthogonal matrix P such that $P\hat{u}_i = \tilde{u}_i$ so that $\Sigma = E(\tilde{u}_i\tilde{u}_i') = I$. P must be of order $T - 1 = 6$ and $PP' = I$. Beyond that we are left with a wide choice of possible P's; we choose Theil's (1971) BLUS procedure as our normalization. In the BLUS procedure P is chosen subject to the two requirements we specified above and to minimize the expected sum of squares of the error vector, where we ignore the β effect which is of order $0_p(NT^{-1/2})$. We define the transformed residuals

$$(2.9) \qquad \tilde{u}_{it} = \hat{u}_{it} - \left(\frac{1}{1 + \sqrt{7}}\right)u_{i1} \qquad \text{for } t=2,\ldots,7$$

We then form the 6 x 6 estimated covariance matrix $\tilde{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} \tilde{u}\tilde{u}'$ which is found in Table 2a. There is still evidence of some serial correlation, positive for adjoining years and negative for years far apart. The likelihood ratio test, for diagonality of the correlation matrix, Anderson (1958, p.233), takes the form

$$(2.10) \qquad \lambda = |\tilde{\Sigma}|^{-N/2}$$

because the diagonal elements of Σ are unity. Since our sample is large, we can use the asymptotic expansion, Anderson (1958, p.239), to calculate $-m \log|\Sigma| = -124.2 \times -1.09 = 135$, for $m = 128 - \frac{23}{6}$ where the second term accounts for higher order terms. Under the null hypothesis this statistic is distributed as central χ^2_{15}. Thus, we reject the null hypothesis of independence even though the deviations exhibited in Table 2, column 6 are not very interesting. Earlier we noted that the data exhibit "overdispersion" which could have led to this rejection even though we conditioned on the firm effect. Since the conditional fixed effects Poisson specification does not solve this problem completely, we may want to allow for another source of within stochastic variation. Such a more general model is considered in the next section.

3. Negative Binomial Models

Even with the fixed effects Poisson model we still have the restriction that the variance and mean are equal, $En_{it} = V(n_{it}) = \lambda_{it}$. On the other hand, the random efects Poisson had a variance to mean ratio of $(\lambda_{it} + \delta)/\delta$ which increases with λ_{it} as our data indicates holds true. Speaking somewhat loosely, we would like to combine the two models to permit the variance to grow with the mean while at the same time we want to have a conditional fixed effect α_i which could be correlated with the right hand side variables, especially R&D. To develop such a model, we begin with the famous negative binomial specification of Yule and Greenwood (1920). We then develop a fixed effects version of the negative binomial specification.

Yule and Greenwood in their model of accident proneness assumed that the number of accidents in a year for a given worker followed a Poisson distribution. They further assumed that the (unconditional) parameter λ_i was distributed in the population randomly and followed a gamma distribution. Our situation differs in two respects from that of Yule and Greenwood. First, we want to specify a conditional model for λ_{it} to ascertain the importance of research and development to the distribution of patents. Also, we have panel data rather than a single cross-section so that we can allow for both the possibility of permanent unobserved firm effects as well as the possibility that these firm effects are correlated with the R&D and other explanatory variables. To start, we return to the situation of Section 1 and consider the yearly patents model. We assume that the Poisson parameter λ_{it} follows a gamma distribution with parameters (γ, δ) and specify $\gamma = e^{X_{it}\beta}$

with δ common both across firms and across time.[1] The mean and variance of λ_{it} are then $E\lambda_{it} = e^{X_{it}\beta}/\delta$ and $V(\lambda_{it}) = e^{X_{it}\beta}/\delta^2$. Note that even if X_{it} remains constant for a firm over time λ_{it} can still vary. This situation should be distinguished from the random effects specification of Section 1 where $\tilde{\lambda}_{it} = \tilde{\alpha}_i e^{X_{it}\beta}$ so that $\tilde{\lambda}_{it}$ was constant for a given firm if the X_{it}'s remained constant. On the other hand, in keeping with the models of Section 2, we have not allowed for firm specific effects. Thus, the λ_{it}'s are independent for a given firm over time.

We now take the gamma distribution for the λ_{it} and integrate by parts to find

$$\text{pr}(n_{it}) = {}_0\!\int^{\infty} \frac{1}{n_{it}!} e^{-\lambda_{it}} \lambda_{it}^{n_{it}} f(\lambda_{it})d\lambda_{it}$$

(3.1)

$$= \frac{\Gamma(\gamma + n_{it})}{\Gamma(\gamma)\Gamma(n_{it} + 1)} \left(\frac{\delta}{1 + \delta}\right)^{\gamma} (1 + \delta)^{-n_{it}}$$

which is the negative binomial distribution with parameters (γ,δ). The log likelihood function is very similar to equation (2.5) with the differences being that here $\gamma = e^{X_{it}\beta}$ rather than being constant and δ is constant rather than being a function of X_{it}. The same comments on computation apply with the use of partial fraction expansions of the gamma and digamma functions permitting rapid evaluation. The moments of n_{it} have the form $En_{it} = e^{X_{it}\beta}/\delta$ and $V(n_{it}) = e^{X_{it}\beta}(1 + \delta)/\delta^2$. Therefore, the variance to mean ratio $V(n_{it})/E(n_{it}) = (1 + \delta)/\delta > 1$. Thus, the negative binomial specification allows for overdispersion with the original Poisson a limiting case as $\delta \to \infty$. We estimate a $\hat{\delta}$ of about .05, implying a variance to mean ratio of 21 which is roughly in line with the evidence in our data that we presented earlier. There are two potential shortcomings of the negative binomial specification: it does not allow for firm specific effects so that serial correlation of the residuals (i.e. non-independence of the counts) may be a problem. Also, the variance to mean ratio is assumed constant across firms while previous evidence demonstrated that the ratio is probably closer to an exponential function of the mean where the exponent moderately exceeds unity.

The estimates from the negative binomial specification are given in the first two columns of Table 3. In the first column we consider a specification with current R&D and time only. The coefficient of R&D equals .75, which is a decrease of 15% from the corresponding model of Table 1 and which may well arise from the overdispersion problem in the original Poisson specification. We also estimate δ to be .04 with quite a small asymptotic standard error. The implied variance to mean ratio is 26, quite far from the original Poisson specification of unity. In column 2 we add the time and R&D interaction along with the scientific sector dummy variable and book value for the firm. The estimated coefficient of current R&D is .56, which is again 15% below the corresponding Poisson model. Time exerts a positive effect while its interaction with R&D has a coefficient of -.012. The declining effectiveness of R&D in producing patents once more seems to be the dominant influence. Lastly, we estimate $\delta = .057$ for a variance to mean

TABLE 3

Estimates of the Negative Binomial Model

	Totals		Random Effects		Fixed Effects	
log R_o	.75 (.02)	.56 (.03)	.52 (.025)	.49 (.04)	.37 (.03)	.42 (.05)
time	-.03 (.01)	.017 (.025)	-.02 (.003)	.01 (.01)	-.02 (.003)	-.005(.010)
time·log R_o		-.012 (.006)		-.010 (.003)		-.004(.003)
dummy (scientific sector)		.40 (.04)		.17 (.11)		.16 (.14)
log book value		.24 (.02)		.13 (.04)		-.06 (.06)
intercept	-1.27 (.07)	-2.20 (.13	1.47 (.10)	.64 (.22)	1.86 (.11)	2.00 (.38)
δ	.04 (.002)	.057 (.003)				
a			2.16 (.30)	2.56 (.35)		
b			1.56 (.31)	1.74 (.37)		
log likelihood	-3,845.3	-3,747.4	-3,310.6	-3,304.9	-2,468.9	-2.467.4
Tests for correlated firm effects					65.0	114.0

TABLE 3a

Estimated Correlation of the Transformed Residuals[1]

	69	70	71	72	73	74
69	1.0					
70	.22	1.0				
71	.18	.28	1.0			
72	.10	.08	.04	1.0		
73	-.003	-.16	-.15	-.04	1.0	
74	.00	-.12	.07	.05	.03	1.0

[1] These residuals are computed using the results in the last column.

ratio of 18.5, which again leads to a strong rejection of the Poisson specification. But as we suspected might happen, when we compute standardized residuals the problem of serial correlation reappears. The correlation matrix is essentially the same as that in Table 1a. Thus, we turn again to a model with firm specific effects to take account of this problem.

In order to add firm specific effects to the negative binomial model we also consider a random effects specification as we did in Section 2 for the Poisson model. It is more convenient in this case, however, first to describe the fixed effects version of our model and then add the random (no correlation with the X's) interpretation to it. To do so we need to find a convenient distribution for the sum of the patents for a given firm $\sum_t n_{it}$ which we will condition on as we did in the Poisson specification of equation (2.5). There once we conditioned on the firm specific effect α_i we returned to a deterministic specification of the λ_{it}. The situation differs here because of the stochastic nature of the λ_{it} even after conditioning. We first find the moment generating function for the negative binomial distribution to be $m(t) = \left(\frac{1 + \delta - e^t}{\delta}\right)^{-\gamma}$. Since the sum of independent random variables equals the product of their moment generating functions we see that if δ is common for two independent negative binomial random variables w_1 and w_2, then $w_1 + w_2 = z$ is distributed as a negative binomial with parameters $(\gamma_1 + \gamma_2, \delta)$. We first derive the distribution, conditioned on z, for the two observation case

$$(3.2) \qquad pr(w_1 | z = w_1 + w_2) = \frac{pr(w_1)pr(z - w_1)}{pr(z)}$$

$$= \frac{\dfrac{\Gamma(\gamma_1 + w_1)}{\Gamma(\gamma_1)\Gamma(w_1 + 1)} (1 + \delta)^{-(w_1 + w_2)} (\frac{\delta}{1 + \delta})^{\gamma_1 + \gamma_2} \dfrac{\Gamma(\gamma_2 + w_2)}{\Gamma(\gamma_2)\Gamma(w_2 + 1)}}{\dfrac{\Gamma(\gamma_1 + \gamma_2 + z)}{\Gamma(\gamma_1 + \gamma_2)\Gamma(z + 1)} (1 + \delta)^{-z} (\frac{\delta}{1 + \delta})^{\gamma_1 + \gamma_2}}$$

$$= \frac{\Gamma(\gamma_1 + w_1)\Gamma(\gamma_2 + w_2)\Gamma(\gamma_1 + \gamma_2)\Gamma(w_1 + w_2 + 1)}{\Gamma(\gamma_1 + \gamma_2 + z)\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(w_1 + 1)\Gamma(w_2 + 1)}$$

Note that in equation (3.2) we are left with the ratio of gamma functions which depend only on the parameter γ, not on the parameter δ. Thus, each firm, in effect, can have its own δ so long as it does not vary over time. The parameter δ has been eliminated by the conditioning arguments.

More generally we consider the joint probability of a given firm's patents conditional on the seven year total

$$(3.3) \qquad pr(n_{i1}, \ldots, n_{iT} | \Sigma n_{it}) = \left(\prod_t \frac{\Gamma(\gamma_{it} + n_{it})}{\Gamma(\gamma_{it})\Gamma(n_{it} + 1)}\right)\left(\frac{\Gamma(\sum_t \gamma_{it})\Gamma(\sum_t n_{it} + 1)}{\Gamma(\sum_t \gamma_{it} + \sum_t n_{it})}\right)$$

The marginal distribution of a given n_{it} conditional on Σn_{it} is a negative hypergeometric distribution (for integer values of the γ_{it}'s) so equation (3.3) is sometimes called a negative multivariate hypergeometric distribution for integer γ_{it}, e.g. Cheng Ping (1964). An alternative derivation instead of proceeding via the negative binomial and conditioning can proceed from the conditional Poisson derivation, i.e. the multinomial distribution of equation (2.6). In equation (2.6) the multinomial parameters arose from the Poisson distribution $p_{it} = \lambda_{it}/\sum_t \lambda_{it} = e^{X_{it}\beta}/\sum_t e^{X_{it}\beta}$. The natural mixing distribution for the multinomial parameters is the Dirichlet distribution which takes the p_{it}'s as random variables on the unit interval and enforces the adding up condition. We then integrate over equation (2.6)

$$(3.4) \quad pr(n_{i1},\ldots,n_{iT}|\Sigma n_{it}) = \frac{(\Sigma n_{it})!}{\prod_t n_{it}!} \int_0^1 \ldots \int_0^1 (\prod_t p_{it}^{n_{it}}) \cdot$$

$$f(p_{i1},\ldots,p_{iT})dp_{i1}\ldots dp_{iT}$$

$$= \frac{(\Sigma n_{it})!}{\prod_t n_{it}!} E(\prod_t p_{it}^{n_{it}})$$

$$= \frac{\Gamma(\sum_t n_{it} + 1)\Gamma(\sum_t \gamma_{it})}{\Gamma(\sum_t \gamma_{it} + \sum_t n_{it})} \prod_t \frac{\Gamma(\gamma_{it} + n_{it})}{\Gamma(\gamma_{it})\Gamma(n_{it} + 1)}$$

where $f(p_{i1},\ldots,p_{iT})$ is the Dirichlet density and $\gamma_{i1},\ldots,\gamma_{iT}$ are the parameters. Note that equations (3.3) and (3.4) are identical as expected. The mean of equation (3.3) is $En_{it} = \gamma_{it}\sum_t n_{it}/\Sigma\gamma_{it}$ which is the same as $\hat{s}_{it}\sum_t n_{it}$ from the multinomial distribution. The variance takes the form of the variance of a multinomial variate times a ratio which arises from the

Dirichlet parameters, $V(n_{it}) = (\sum_t n_{it}\gamma_{it}/\Sigma\gamma_{it})(1 - \gamma_{it}/\Sigma\gamma_{it})(\Sigma n_{it} + \Sigma\gamma_{it})/(1 + \Sigma\gamma_{it})$. We have, thus, again increased the variance over the multinomial case (equation 2.2) to allow for overdispersion. Now the Dirichlet distribution occurs because each λ_{it} is distributed as a gamma random variable with parameters (γ_{it},δ). Rescale the variables to take the form $(\delta^{-1}\gamma_{it},1)$ and note that the random variable $\lambda_{i1}/(\lambda_{i1} + \lambda_{i2})$ is distributed as a beta random variable variable with parameters $(\delta^{-1}\gamma_{i1},\delta^{-1}\gamma_{i2})$. The Dirichlet distribution is the multivariate generalization of the beta distribution. Thus the random vector $(\lambda_{i1}/\sum_t\lambda_{it},\ldots,\lambda_{iT}/\sum_t\lambda_{iT})$ is distributed as a Dirichlet random vector with parameters Θ_{it}, $t=1,\ldots,T$ for $\Theta_{it}=\delta^{-1}\gamma_{it}$.[1] We have derived the conditional negative binomial model in two ways: The first finds the conditional model for the negative binomial specification. Equivalently, one can begin with the conditional Poisson model and let the λ_{it}'s be random variables. Both derivations yield interesting insights into the basic model.

The log likelihood of the sample follows once we specify γ_{it}. We let the parameters of the underlying model be $(\gamma_{it}, \delta_i) = (e^{X_{it}\beta}, \phi_i/e^{\mu_i})$where both ϕ_i and μ_i are allowed to vary across firms. The mean of $\lambda_{it}=\left(e^{X_{it}\beta + \mu_i}\right)/\ \phi_i$ while the variance is $V(\lambda_{it}) = \left(e^{X_{it}\beta +2\mu_i}\right)/\phi_i^2$. Therefore, we have multiplied the mean by e^{μ_i} as we did for the deterministic Poisson parameter in the fixed effects case. Likewise, the standard deviation has been multiplied by the same amount. Considering the unconditional negative binomial model we calculate $En_{it} = (e^{X_{it}\beta + \mu_i})/\phi_i$ with $V(n_{it}) = (e^{X_{it}\beta + \mu_i}/\phi_i^2)\,(1 + \phi_i/e^{\mu_i})$ so that the variance to mean ratio is $(e^{\mu_i}/\phi_i)(1+\phi_i/e^{\mu_i})$. Thus we allow for both overdispersion, which the fixed effects Poisson specification did not, as well as a firm specific variance to mean ratio, which the original negative binomial specification did not. With this specification of γ_{it} we find the log likelihood for the conditional fixed effects specification of equation (3.3)

$$(3.5) \qquad L(\beta) = C_4 + \sum_i [\log \Gamma(\sum_t e^{X_{it}\beta}) - \log \Gamma(\sum_t e^{X_{it}\beta} + \sum_t n_{it})]$$
$$+ \sum_i\sum_t [\log \Gamma(e^{X_{it}\beta} + n_{it}) - \log\Gamma(e^{X_{it}\beta})]$$

Estimates for the fixed effects negative binomial model are given in columns 5 and 6 of Table 3. The first set of coefficient estimates are quite close to the conditional Poisson model with the coefficient of R&D about one-half as large as the original negative binomial specification. When we interact time and R&D in column 6 we do find important differences from the Poisson fixed effects model. First, the estimate of the coefficient of current R&D is .42, which is somewhat lower than the Poisson model estimate of .48. Next the pure time effect continues to be negative, although insignificantly so, while in all previous models it becomes positive when the interaction term was added. Correspondingly, the interaction term has a much smaller estimated magnitude. This last set of results continues to indicate the decline in effectiveness of R&D in producing patents. But at the midpoint of our time period the pure time effect was almost of the same size as the interaction term. Thus, factors beyond the effectiveness of R&D may have been important in explaining the decline in patenting.

As a check, we again compute residuals for the last fixed effects specification. We use equation (2.8) to transform the residuals to approximate independence under the null hypothesis of no serial correlation and estimate the 6 x 6 corelation matrix which is given in Table 3a. There is no sign of serious serial correlation now. The likelihood ratio test based on equation (2.9) is calculated as $-m\log|\tilde{\Sigma}| = 28.7$, which under the null hypothesis is distributed as central χ^2_{15}. It barely rejects the null hypothesis of independence at the five percent level. The fixed effects specification comes close to solving the non-independence problem.

In the 1940's Feller (1943) noticed the similarity of the mathematical forms of the Yule-Greenwood (1920) overdispersion model and the Polya-Eggenberger (1923) model of contagion. He drew the distinction between <u>true contagion</u> (Polya-Eggenberger), where every favorable event increases (or decreases) the probability of future favorable events, and <u>apparent contagion</u> (Yule-Greenwood), which arises from population heterogeneity. In a single cross-section the models cannot be distinguished as Feller points out (p.398), "an excellent fit of Polya's distribution to observations is not necessarily indicative of any phenomenon of contagion in the mechanism behind the observed distribution." With patents we might well expect contagion to be present. One good idea might lead to several patents. Alternatively, the "stock of knowledge leading to patents" may be depleted so that future patents, conditioned on R&D, may be less likely. Although Feller is correct about the cross-section case, can we look for true contagion in our panel data? The answer seems to be yes. Population heterogeneity is accounted for by the presence of the fixed effect after taking account of differences in R&D. True contagion should then appear as either positive serial correlation or negative serial correlation in the residuals. We find little evidence of such serial correlation empirically. Note, however, that our evidence is not as compelling as it might be. Our observation interval is a year, which may tend to obscure the presence of contagion if it occurs for short periods of time relative to one year's time. In order to be able to give a more definitive answer, observations on the arrival time of patents within a year would be required to test for "bunching up" of arrivals. We would then require a different sampling scheme than we currently have to make a study of the correlation of the arrival time intervals. But this type of data is rather artificial within the institutional and legal process for patent applications. Also, the assumption of a fixed amount, i.e. uniform flow, of R&D during the year would need to be modified. Attempting to investigate the finer pattern of the data to probe the contagion question does not appear promising. But looked at on a yearly basis, the presence of firm specific effects together with the absence of substantial serial correlation does indicate the absence of contagion "in the large." It is instructive to note how the presence of serial correlation in the original Poisson or negative binomial model without firm specific effects might mislead us into incorrectly concluding the opposite, much in line with Feller's remarks.

We lastly consider the random effects version of the negative binomial specification. In the fixed effects specification we set the parameters of the underlying model as $(\gamma_{it}, \delta_i) = (e^{X_{it}\beta}, \phi_i/e^{\mu_i})$ so that both ϕ_i and μ_i vary across firms. Upon conditioning on the total number of patents in equation (3.4), the ϕ_i and μ_i parameters are eliminated and only $\gamma_{it} = e^{X_{it}\beta}$ appears. Analogously to the Poisson random effects specification, we now assume that ϕ_i and μ_i are randomly distributed across firms, independent of the X_{it}'s. Given the choice of a probability distribution, we then have the random effects negative binomial specification. An interesting difference exists between the Poisson random effects specification and the negative binomial random effects specification. In the Poisson case, $\tilde{\lambda}_{it} = \lambda_{it}\tilde{\alpha}_i$ where $\tilde{\alpha}_i$ is a random firm specific effect. Note that for constant λ_{it}, $\tilde{\lambda}_{it}$ is also constant, which would occur if the X_{it}'s are constant. However, in the negative binomial specification λ_{it} varies randomly across years even if the X_{it}'s are constant because it is a realization from a gamma probability distribution each year. Thus, we have randomness both across firms and across time, which corresponds to the usual specification in the linear case where we have the variance components decomposition for the stochastic disturbance $\varepsilon_{it} = \alpha_i + \eta_{it}$.

We now choose a distribution for e^{μ_i} and for ϕ_i which will allow us to integrate δ_i out of the marginal probability statement

(3.6) $\text{pr}(n_{i1},\ldots,n_{iT}|X_{i1},\ldots,X_{iT}) = p(n_{i1},\ldots,n_{iT}|X_{i1},\ldots,X_{iT},\mu_i,\phi_i)g(\mu_i,\phi_i)$

where $g(\cdot)$ is the probability density of the incidental parameters. Again, let us assume that e^{μ_i} is distributed as a gamma random variable with parameter (b, 1). Furthermore, we also assume that ϕ_i is distributed as an independent gamma random variable with parameters (a, 1)[3]. Consider the ratio $\delta_i/(1 + \delta_i) = \phi_i/(e^{\mu_i} + \phi_i)$ which can then be shown to be distributed as a beta random variable with parameters (a, b). Therefore, $\delta_i/(1 + \delta_i)$ has a density function $f(Z) = [B(a, b)]^{-1} z^{a-1}(1 - z)^{b-1}$ where $B(\cdot)$ is the beta function. The ratio $\delta_i/(1 + \delta_i)$ takes values on the unit interval which is appropriate for $\delta_i > 0$. The mean is $E(\delta_i/(1 + \delta_i)) = a/(a + b)$ with variance $V(\delta_i/(1 + \delta_i)) = ab/(a + b + 1)(a + b)^2$. We integrate using the beta density to find

$$\text{(3.7)}\quad \text{pr}(n_{i1},\ldots,n_{iT}|X_{i1},\ldots,X_{iT})$$

$$= \int_0^1 \prod_{i=1}^{T} \left[\frac{\Gamma(\gamma_{it} + n_{it})}{\Gamma(\gamma_{it})\Gamma(n_{it} + 1)} z_i^{\gamma_{it}} (1 - z_i)^{n_{it}} \right] f(z_i)dz_i$$

$$= \frac{\Gamma(a + b)\Gamma(a + \Sigma\gamma_{it})\Gamma(b + \Sigma n_{it})}{\Gamma(a)\Gamma(b)\Gamma(a + b + \Sigma\gamma_{it} + \Sigma n_{it})} \prod_t \frac{\Gamma(\gamma_{it} + n_{it})}{\Gamma(\gamma_{it})\Gamma(n_{it} + 1)}$$

where $z_i = \delta_i/(1 + \delta_i)$. Note that the last term in equation (3.7) corresponds exactly to a term in the fixed effects model of equation (3.4). But we now estimate additional parameters a and b from the beta distribution which describe the distribution of the δ_i across firms. The log likelihood function has the form

$$(3.8) \quad L(\beta,a,b) = C_5 + N \log[\Gamma(a + b)/\Gamma(a)\Gamma(b)] + \sum_i \log \Gamma(a + \sum_t e^{X_{it}\beta})$$

$$+ \log \Gamma(b + \sum_t n_{it}) - \log \Gamma(a + b + \sum_t e^{X_{it}\beta} + \sum_t n_{it})]$$

$$+ \sum_i \sum_t [\log \Gamma(e^{X_{it}\beta} + n_{it}) - \log \Gamma(e^{X_{it}\beta})].$$

where C_5 is a constant. Estimates of the unknown parameters are given in columns 3 and 4 of Table 3.

The results of the random effects negative binomial specification fall in between the estimates from the totals model and the estimates from the fixed effects model. In column 3 of Table 3, where only R&D and time are used in the specification for γ_{it}, the coefficient of R&D is estimated to be .52, compared to .75 for the totals model and .37 for the fixed effects model. The estimate of the time coefficient is negative and the same as the fixed effects estimate. The parameters of the beta distribution are estimated quite precisely along with a large increase in the likelihood function compared to the totals model. The estimated mean δ is 1.38, which is significantly higher than for the totals model where the estimate of (fixed) δ is .04. The variance to mean ratio (at the mean) is now estimated to be 1.72. But now the variance to mean ratio is allowed to vary across firms rather than taking on a constant value as it does in the totals model. A Hausman test of the random versus fixed effects specification yields 65.0 which leads to a rejection of the hypothesis of no correlation between the δ_i and R&D. This result was to be expected, given the evidence in Figure 1 that δ_i is negatively correlated with R&D.

In column 4 of Table 3 we now include the R&D-time interaction term and the two firm specific variables, book value and scientific sector. The results differ markedly from the Poisson case where this specification gave almost identical results for the random effects and fixed effects models. Here the estimates of the coefficients of R&D and book value differ significantly in the two cases. The Hausman test statistic equals 114.0, which clearly rejects the no correlation hypothesis. But the reason for the rejection becomes apparent when we consider the differences between the random effects Poisson specification and the random effects negative binomial specification. In the latter, we estimate the variance to mean ratio to be 1.68. However, the negative correlation between R&D and δ_i remains, which leads to the rejection using the Hausman test. In the corresponding random effects Poisson specification, δ_i is not estimated (since the variance to mean ratio is set to one) and only a firm specific mean effect μ_i is allowed to vary across firms. When both book value and the scientific sector dummy are included in the Poisson model, they "explain" the firm effect. Therefore, the Hausman test does not reject the no correlation assumption. In the random effects negative binomial specification where both μ_i and δ_i are permitted to vary, inclusion of book value and the scientific sector dummy do not remove the correlation between δ_i and R&D. Thus, the Hausman test rejects the no correlation assumption of the random effects specification in favor of the fixed effects specification, which conditions on individual firm values of μ_i and δ_i without needing the no correlation assumption.

4. Between Firm Models

Within the context of the linear panel data models it is often useful to separate the total sample variability into between firm and within firm variability. That is, given the model $y_{it} = X_{it}\beta + \alpha_i + \eta_{it}$, $i=1,N$ and $t=1,T$, the between model takes the form $y_{i.} = X_{i.}\beta + \alpha_i + \eta_{i.}$ where the dot notation signifies time averages, e.g. $y_{i.} = \frac{1}{T}\Sigma y_{it.}$ The corresponding within model is given by $(y_{it} - y_{i.}) = (X_{it}-X_{i.})\beta + \eta_{it} - \eta_{i.}$. This decomposition is unique and the resulting samples are orthogonal which can easily be seen by noting that if we stack the original model $y = X\beta + \varepsilon$ for $\varepsilon_{it} = \alpha_i + \eta_{it}$ the between model arises from the projection $P_e y = P_e X\beta + P_e\varepsilon$ where $P = I \otimes e(e'e)^{-1}e'$, e a T long vector of ones. The within specification arises from the orthogonal projection $Q_e y = Q_e X\beta + Q_e\varepsilon$ where $Q_e = I \otimes (I - P_e)$. Both models are of interest although often one cannot obtain unbiased estimates of the parameters in the between model because of lack of independence between the X_{it} and the firm specific effect $\alpha_{i.}$[1] Now the conditional models we have been considering are analogous to the within models for the linear specification. We demonstrate this fact by assuming that α_i and η_{it} are normally distributed, and conditioning on the mean number of patents

$$(4.1) \qquad E(y_{it}|y_{i.}) = X_{it}\beta + \frac{\mathrm{cov}(y_{it},y_{i.})}{\mathrm{var}(y_{i.})}(y_{i.} - X_{i.}\beta)$$

$$= X_{it}\beta + y_{i.} - X_{i.}\beta$$

which, rearranged to $E(y_{it} - y_{i.}) = (X_{it} - X_{i.})\beta$, yields the within specification. However, in our models we cannot use linear projections which separate the variables uniquely into $x_{i\cdot}$ and $x_{it} - x_{i\cdot}$ components. We explore the parallel definition of "between" models in this section. Our first conditional model, the fixed effects Poisson specification, separates the original total sample into a conditional multinomial probability times a marginal Poisson probability

$$(4.2) \qquad pr(n_{i1},\ldots,n_{iT}|X_{i1},\ldots,X_{iT})$$

$$= pr(n_{i1},\ldots,n_{iT}|X_{i1},\ldots,X_{iT},\sum_t n_{it},\alpha_i)\ pr(\sum_t n_{it},\alpha_i|X_{i1},\ldots,X_{iT}).$$

The first probability of the right hand side of equation (4.2) was derived in equation (2.5) to be a multinomial distribution. The marginal probability follows from taking the product of the moment generating function of the Poisson distribution $\prod_{s=1}^{T} m_s(t) = \prod_s e^{-\sum_s \lambda_s} e^{\sum_s \lambda_s e^t}$ so that the sum $\sum_t n_{it}$ is distributed as Poisson with parameter $\Lambda_i = \sum_s \lambda_{is} = T\lambda_{i.}$. We need to integrate out the unobservable random firm effect α_i from the marginal probability for $\sum_t n_{it}$ in equation (4.2). Therefore as we did in equation (2.3) we assume that $\alpha_i = e^{\mu_i}$ is distributed as a gamma random variable with parameters γ and δ. We use the results of equation (2.3) on the sum of the patents $\sum_t n_{it}$ to derive the marginal probability

$$(4.3) \qquad \mathrm{pr}(\sum_t n_{it} | X_{i1},\ldots,X_{iT})$$

$$= \left(\sum_t e^{X_{it}\beta}\right)^{\sum_t n_{it}} \left[\frac{\delta}{\sum_t e^{X_{it}\beta} + \delta}\right]^{\gamma} \left[\left(\sum_t e^{X_{it}\beta} + \delta\right)\right]^{-\sum_t n_{it}} \frac{\Gamma(\gamma + \sum_t n_{it})}{\Gamma(\gamma)\Gamma(\sum n_{it} + 1)}$$

Note that as with the linear between specification, the between Poisson model suffers from the same problem as the random effects Poisson specification -- it assumes that the firm specific effects are uncorrelated with the explanatory variables, including R&D. Note also that all the X_{it} enter the between model in equation (4.3) instead of just $X_{i\cdot}$ appearing.

The log likelihood for the between model is written

$$(4.4) \qquad L(\beta, \gamma, \delta) = C_6 - N\log\Gamma(\gamma) + N\gamma\log\delta + \sum_{i=1}^{N} [\Sigma n_{it}\left(\log(\sum_t e^{X_{it}\beta}) - \right.$$

$$\left.\log(\sum_t e^{X_{it}\beta} + \delta)\right) - \gamma\log(\sum_t e^{X_{it}\beta} + \delta) + \log\Gamma(\gamma + \sum_t n_{it})]$$

where C_6 is a constant. Because of the nonlinearity introduced by the exponential functions the between model does not depend on $X_{i\cdot}$ (or $TX_{i\cdot}$) like the linear between model but instead depends on the within period variation of the X_{it} via $\sum_t e^{X_{it}\beta}$. Still, a close relationship to the linear case exists. Rather than partitioning the sums of squares into a between and within component, we partition the likelihood of the original sample into two components, conditional and marginal, so that the log likelihoods add up L(β, nil,...,niT) = LC(β, nil,...,niT| Enit) + LM(β, Enit) for a common parameter vector β. The log likelihood function on the left hand side of the equation is given by equation (1.2) while the conditional log likelihood $L_C(\cdot)$ is in equation (2.5) and the marginal log likelihood L_M is the between model of equation (4.4). Similarly, the Fisher information regarding the parameters add up, $J_T = J_C + J_M$ for $J_T = -\lim E\frac{\partial^2 L}{\partial\beta\partial\beta'}$ with the variance matrices for the estimates β following by matrix inversion. Although the interpretation is not as neat in the Poisson case as in the linear case where no within sample variation enters the between model, the idea of partitioning the information in the data into two additive components still goes through.

In the first two columns of Table 4 we give the estimates of the between Poisson specification. The coefficient of current R&D expenditures is somewhat less than that of the original Poisson model. The estimates of γ and δ imply a ratio of variance to mean for the firm effect of about 4 rather than 10, which we obtained for the random effects model on individual years of data. When we add the firm variables, however, this ratio again becomes near unity as in Table 2. The size of the coefficient on time interacted with R&D suggests that the earlier R&D expenditures are substantially more important than the later expenditures for the overall level of patents; this result is consistent with the U-shaped lag we saw in the random effects estimates in Table 2.

A similar decomposition of the original negative binomial specification of equation (3.1) exists. The within specification is given in equation (3.3) as a ratio of gamma functions which takes the form of a negative

TABLE 4

Estimates of Marginal ("Between") Firm Models

	Poisson		Negative Binomial	
log R_o	.75 (.04)	1.18 (.15)	.79 (.05)	1.27 (.45)
time	--	--	-.22 (.16)	.67 (.49)
time·log R_o		-.26 (.04)		-.15 (.07)
dummy (scientific sector)		.20 (.18)		.39 (.10)
log book value		.29 (.10)		.23 (.05)
intercept			-2.07 (.55)	-7.08 (2.93)
γ	1.29 (.15)	1.55 (.17)		
δ	.23 (.03)	.87 (.03)		
a			9.1 (4.4)	151 (694)
b			731.7 (470)	15154 (71340)
log likelihood	-806.5	-792.9	-790.0	-776.0

multivariate hypergeometric distribution. The between negative binomial model follows from taking the product of the moment generating function of the negative binomial distribution under the assumption that the underlying gamma distribution for the Poisson parameter λ_{it} has γ_{it} varying across years but keeps δ_i constant. The moment generating function is

$$(4.5) \qquad \prod_{s=1}^{T} m_s(t) = \left(\frac{1 + \delta_i - e^t}{\delta_i}\right)^{-\sum_s \gamma_{is}}$$

which is a negative binomial with parameters $(\sum_t \gamma_{it}, \delta_i)$. To derive the between firm negative binomial model we make the same assumptions about δ_i as we did earlier for the derivation of the random effects model (equation 3.7).

We take the negative binomial distribution with parameters $(\sum_t \gamma_{it}, \delta_i)$ and specify $\delta_i/(1 + \delta_i)$ to be distributed as a beta random variable so that the between firm specification takes a generalized hypergeometric form,

$$(4.6) \qquad pr(\Sigma n_{it} | X_{i1},\ldots,X_{iT}) =$$

$$\frac{\Gamma(\Sigma\gamma_{it} + \Sigma n_{it})\ \Gamma(a + b)\ \Gamma(a + \Sigma\gamma_{it})\ \Gamma(b + \Sigma n_{it})}{\Gamma(\Sigma\gamma_{it})\ \Gamma(\Sigma n_{it} + 1)\ \Gamma(a)\ \Gamma(b)\ \Gamma(a + b + \Sigma\gamma_{it} + \Sigma n_{it})}$$

where a and b are the parameters of the underlying beta distribution. The log likelihood function for equation (4.6) follows directly. It is interesting to note that in equation (4.6) the leading terms in the numerator and denominator arise from the combinatorial term in the negative binomial distribution of equation (2.5) while the remaining terms arise from the ratio of two beta functions.

In columns 3 and 4 of Table 4 we give the results of the between negative binomial model of equation (4.6). The most striking difference between these estimates and those of the between Poisson model is that the negative effect of later R&D expenditures has become a small positive effect; the steep slope of the R&D coefficient has flattened and the overall level is much higher (.67 instead of .14 for constant R&D). Both models give the not very surprising result that firm characteristics such as industry and size are as important as R&D expenditures for the overall level of firm patenting. These results should not be taken too seriously, however, since there is very little relevant variance in the marginal (between) dimension of the data to identify the fine time structure of R&D effects.

5. Summary

Our various models can be thought of as differing along two conceptual dimensions: (1) where and to what extent do they allow for "disturbances in the equation," for variability not explicitly accounted for either by the X's or by the assumed underlying Poisson process, and (2) are the relevant coefficents (β's) different when estimated in the conditionial ("within") rather than in the marginal ("between") dimension of the data. That is, do we get different answers when we focus on the shorter term time-series aspects of the data than when we sum or average over a longer time period and use primarily the cross-sectional aspect of the data. In Mundlak's (1978) language, are the individual "effects" correlated with the X's?

Table 5 attempts to organize and summarize all of our different models. We start with the "total" Poisson: It assumes no disturbances in the equation and maintains the equality of coefficients across all dimensions of the data. It can be partitioned into two components: conditional ("within") and a marginal ("between"). If the two yielded the same estimated coefficients, their log likelihoods would sum to the earlier total. The actual sum is higher, implying that the coefficients do differ (as can also be seen in column 3), that there is a correlation between individual firm effects and their R&D expenditures.

All the other models represent different ways of adding randomness. The Poisson "random effects" model adds a pure firm disturbance with no within (year to year) variability. Note the large increase in the log likelihood (from -9,078 to -3,780). The Negative Binomial "total" allows the Poisson

TABLE 5

Summary of Results

Model	Log Likelihood Poisson	Log Likelihood Negative Binomial	Total R&D Coefficient[2] Poisson	Total R&D Coefficient[2] Negative Binomial
1. Totals (no firm effects)	-9,077.5	-3,747.4	.57 (.006)	.51 (.02)
2. Marginal (no firm effects)	-6,065.2	-776.1	.56 (.008)	.66 (.19)
3. Conditional	-2,979.0	-2,467.4	.41 (.03)	.40 (.04)
Sum of 2 and 3	-9,044.2	-3,243.5		
Test of 2 and 3[1] against 1	$\chi_2^2=66.6$	$\chi_6^2=1008.$		
4. Totals (random effects)	-3,779.6	-3,304.9	.41 (.01)	.45 (.04)
5. Marginal (random effects)	-792.9	-776.0	.14 (.13)	.67 (.19)
Sum of 5 and 3	-3,771.9	-3,243.4		
Test of 5 and 3[1] against 4	$\chi_2^2=15.4$	$\chi_6^2=123.$		

[1] These tests are likelihood ratio tests for the equality of the coefficients in the marginal and conditional models.

[2] This coefficient is computed as the total effect of log R&D in 1971, $\beta_R + 4\cdot\beta_{t\cdot R}$.

parameter λ_{it} to be distributed randomly, across firms and time, according to a Gamma distribution. Adding such a disturbance again increases the likelihood greatly (from -9,078 to -3,747). The random effects negative binomial, which is in effect a Beta distribution (as described in the previous section), allows the variance of the effects to differ in the within and between dimensions. It is essentially a "variance components" version of the negative binomial. It is clear from the results reported in Table 5 that the data want both a disturbance in the conditional within dimension (compare the conditionals for the negative binomial and Poisson) and a different one, with a different variance, in the marginal (between) dimension. The big changes in fit come from the introduction of such variability and from allowing it to differ across these two dimensions of the data.[1] Most of this variability is in the between dimension (compare the log likelihoods for the two Poisson marginals, one without and the other with firm effects), but there is also variability in the time dimension. The estimated coefficients differ in the two dimensions, but much less so (the likelihood rises only from -3,305 to -3,245).

Substantively, our results differ from those of Pakes and Griliches (1980) primarily because of the introduction of additional firm specific variables (log book value and scientific industry dummy) and the log R-time interaction. Adding the firm specific variables reduces the coefficient of log R from about .8 to .6 and brings the "between" and "within" estimates closer to each other. While there is still some (positive) correlation left between the individual firm propensity to patent and its R&D intensity, it is now much smaller. In fact, it would not be a bad approximation to assume that controlling for industry and size, the remaining firm effects are largely random.

Our estimate of the elasticity of patenting with respect to R&D is about .4 and slightly lower than those of Pakes and Griliches. That is partly due to our inclusion of only one log R term in the final analysis. The data we used were not informative enough about the lag structure, and considering the already complex nature of our computations, we did not experiment much with this aspect of the specification. A more adequate specification of the lag structure might have raised the sum of estimated coefficients by .1 or .2.

The major new substantive finding is that the negative trend in the patent data has a strong interactive component. That is, rather than the propensity to patent just declining exogenously over time, firms are getting less patents from their more recent R&D investments, implying a decline in the "effectiveness" or productivity of R&D.

Methodologically, we have shown how a panel of count data can be analyzed consistently.[2] We described and illustrated the theoretical and empirical necessity to generalize the Poisson model to allow for both "individual" effects and for "overdispersion" in the data and derived models which allowed us to do it. More work needs to be done, however, on the analysis of residuals from such models. Also, it would be interesting to introduce firm effects which could decay over time. This would allow us to consider the effects of lag truncation in such models (along the lines of the Griliches-Pakes work for linear distributed lag models). But even without such refinements, this type of model has many potential uses in econometric data analysis which we expect to pursue further in the future.

NOTES

Section 1

[1]It has a long history in the analysis of accident data with perhaps the most famous example being von Bortkiewicz's 1898 study of accidental death by mule kick in the German army.

[2]It is interesting to note how similar our results are to the OLS results which are unbiased (except for the zero problem) in the presence of serial correlation. It may turn out to be the case that the main effect of serial correlation is on the estimated standard errors which are inconsistent for both estimators.

Section 2

[1]This problem has been recently discussed by Mundlak (1978), Hausman (1978), Chamberlain (1980), and Hausman-Taylor (1980).

[2]Note that this specification is close to the classic Yule-Greenwood (1920) specification which leads to a negative binomial specification. A similar probability specification was derived by Bates and Neyman (1952) for a somewhat different model of accident proneness from that of Yule and Greenwood. Bates and Neyman named the distribution the multivariate negative binomial distribution. It is also referred to as the negative multinomial distribution.

[3]Chamberlain (1980) also derives a multinomial logit in his generalization of Cox's (1970) fixed effects binomial logit model.

[4]With more years of data we might well want to let this initial stock of knowledge decay over time. However, we did not find evidence of such a decay process in our residuals.

[5]It may be interesting to report also the comparable original OLS estimates for this model. Without the time interaction and firm specific variables the estimated coefficient of log R is .81, .77, .29, and .39 for the total, within, between, and variance-components specifications respectively. With the additional variables they are .49, .54, .29, and .29. The variance-components results are close to the within because most of our variance is between (95 percent for log Patents and 97 percent for log R) which is downweighted in this specification. These results are mirrored in the random-effects specification results reported in the text. Note, however, that the comparable results are somewhat higher for the Poisson than the OLS specification.

Section 3

[1]We could easily specify δ to change over time. However, preliminary investigation did not demonstrate that this more general specification was needed.

[2]Note that the scale parameter δ is not identified here. We set $\delta=1$. This result is to be expected for the conditional model given the results of equations (3.2) and (3.3).

[3]Since these are unobservable random variables, the scale parameter merely serves as a normalization.

Section 4

[1]Hausman (1978) considers a test for the presence of this correlation which we have used in the preceding sections. Hausman-Taylor (1980) further investigate the problem and devise a technique for consistent estimation, even when correlation is present.

Section 5

[1]The log likelihood of the marginal negative binomial with random firm effects is about the same without such effects, indicating that once the data are summed, one disturbance is enough, the model cannot distinguish between two different sources of variance.

[2]After this paper was written, an unpublished paper by G.C. Gilbert, "Econometric Models for Discrete Economic Processes," which covers some of of the same topics, was brought to our attention. Our basic model is the same as his "multiplicative" form of the Poisson model. However, the present paper extends the Poisson and Negative Bionomial models to the panel data setting where the independence of each observation is not a reasonable assumption. In addition, we have obtained all our maximum likelihood estimates by unconstrained optimization of the log likelihood function and did not find the problems in obtaining estimates of the negative binomial model that he found.

REFERENCES

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

Andersen, E.B. (1972), "The Numerical Solution of a Set of Conditional Estimation Equations," Journal of the Royal Statistical Society, B, 34.

Andersen, E.B. (1973), Conditional Inference and Models for Measuring, (Copenhagen).

Anderson, T. (1958), Multivariate Statistical Analysis, (New York).

Bates, G.E. and J. Neyman (1952), "Contributions to the Theory of Accident Proneness," University of California Publications in Statistics, 1.

Chamberlain, G. (1980), "Analysis of Variance with Qualitative Data," Review of Economic Studies, 47.

Cheng Ping (1964), "Minimax Estimates of Parameters of Distributions Belonging the Exponential Family," Acta Mathematica Sinica, 5.

Chernoff, H. and E. Lehmann (1954), "The Use of the Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit," Annals of Mathematical Statistics, 25.

Cox, D.R. (1970), Analysis of Binary Data, (London).

Cox, D.R. and E.J. Snell (1968), "A General Definition of Residuals," Journal of the Royal Statistical Society, B, 30.

Eggenberger, F. and G. Polya (1923), "Über die Statistik verketteter Vorgänge," Zeitschrift fur Angewandte Mathematik und Mechanik, 1.

Feller, W. (1943), "On a General Class of Contagious Distribution," Annals of Mathematical Statistics, 14.

Gilbert, G.C. (1979), "Econometric Models for Discrete Economic Processes," given at the European Meetings of the Econometric Society, Athens, 1979.

Greenwood, M. and G.U. Yule (1920), "An Inquiry into the Nature of Frequency Distribution of Multiple Happenings," Journal of the Royal Statistical Society, A.

Pakes A. and Z. Griliches (1980), "Estimation of Distributed Lags in Short Panels, NBER Working Paper #4.

Haberman, S. (1977), "Maximum Likelihood Estimates in Experimental Response Models, Annals of Statistics, 5.

Hausman, J.A. (1978), "Specification Tests in Econometrics," Econometrica, 46.

Hausman, J.A. and W.E. Taylor (1980), "Panel Data and Unobservable Individual Effects," forthcoming in Econometrica.

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," Econometrica, 46.

Neymann, J. and E.L. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations, Econometrica, 16.

Pakes, A. and Z. Griliches (1980), "Patents and R&D at the Firm Level: A First Look," NBER Working Paper.

Rao, C.R. (1952), Advanced Statistical Methods in Biometric Research, (John Wiley: New York).

Theil, H. (1971), Principles of Econometrics, (John Wiley: New York).