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

By Jerry Hausman, Bronwyn H. Hall, and Zvi Griliches[1]

This paper focuses on developing and adapting statistical models of counts (nonnegative integers) in the context of panel data and using them to analyze the relationship between patents and R & D expenditures. Since a variety of other economic data come in the form of repeated counts of some individual actions or events, the methodology should have wide applications.

The statistical models we develop are applications and generalizations of the Poisson distribution. Two important issues are (i) Given the panel nature of our data, how can we allow for separate persistent individual (fixed or random) effects? (ii) 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. Lastly, we develop a test statistic for the presence of serial correlation when fixed effects estimators are used in nonlinear conditional models.

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. (i) 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. (ii) 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 (nonnegative 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 spells of sickness in a year, the number of records purchased per month, the number of cars owned, or the number of jobs held during a year, all have nonnegligible probabilities of zero and are nonnegative 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. (i) Given the panel nature of our data, how can we allow for separate persistent individual (fixed or random) effects? (ii) How does one introduce the equivalent

of disturbance-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 [25]. In that work patent data for 8 years (1968–1975) and 121 U.S. companies were analyzed as a 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 per cent 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.[2] 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 per cent). 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 Compustat tape and other sources (see Pakes and Griliches [25] for more detail on sample derivation and construction), and deflated by an approximate R & D cost deflator.

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

---

[2] It is difficult to distinguish in a short 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 [15] for further discussion of these issues.

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 a random distribution of its own, leading to the estimation of a negative-binomial model for these data. Section 4 explores in our nonlinear 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.[3] 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: (i) In many ways it is analogous to the familiar econometric regression specification. In particular, $E(n_{it} \mid X_{it}) = \lambda_{it}$. 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. (ii) 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. (iii) 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 one to T are also Poisson with parameter $\lambda_i = \sum_{t=1}^{T} \lambda_{it}$. 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

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

---

[3] 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. The Poisson and subsequent models that we consider might also usefully be analyzed as members of the "generalized linear model" class of Nelder and Wedderburn [22]. See also Johnson and Kotz [19]. Gilbert [12] has applied the Poisson model to economic data.

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 comes 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 this Poisson specification is

$$(1.2) \qquad L(\beta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[n_{it}! - e^{X_{it}\beta} + n_{it}X_{it}\beta \right].$$

The gradient and Hessian take the forms

$$(1.3) \qquad \frac{\partial L}{\partial \beta} = \sum \sum \left[X_{it}'(n_{it} - e^{X_{it}\beta}) \right],$$

$$\frac{\partial^2 L}{\partial \beta \partial \beta'} = \sum \sum \left[-(X_{it}'X_{it})e^{X_{it}\beta} \right].$$

The first order conditions indicate that β can be estimated either by an iterative nonlinear weighted least squares program with $n_{it} - \lambda_{it}$ as the "residual" or by a maximum likelihood (ML) program. The Hessian demonstrates that the likelihood function is globally concave as 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 $\hat{\beta}$.

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 I. We also present the corresponding estimates of a least squares regression of $\log(n_{it}) = X_{it}\beta + \epsilon_{it}$ 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 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 cent per year. Lastly, we have the somewhat disturbing pattern of a U-shaped distributed lag which may well indicate a substantial truncation effect. This pattern disappears, however, when we allow for firm specific effects below.

We now consider alternative specifications of the basic Poisson model. In column 4 of Table I we include only contemporaneous R & D since we find later in the paper that when firm specific effects are added the lagged effects become quite small and difficult to identify. Note that the coefficient of current R & D is very close to the sum of the coefficients in our initial specification. The ex-

TABLE I

ESTIMATES OF THE PATENTS MODEL—TOTALS[1]

Variable	Mean	Coefficients				
		Poisson	OLS[2]	Poisson	OLS	Poisson
$\log R_0$	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 (0.15)	.04 (.01)
time × $\log R_0$						−.02 (.008)
dummy[3] (scientific sector)	0.52					.37 (.01)
log book value[4]	5.36 (2.00)					.24 (.004)
intercept		1.68 (0.20)	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 of residuals			.869		.914	
log likelihood		−10,621.1		−11,198.4		−9,077.5

[1] The sample is 128 firms, annual data from 1968 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 electric component industries.

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

[5] Sum evaluated at midpoint of the period, 1971.

ogeneous time effect has now decreased in magnitude to 4 per cent per year. In column 5 we find very similar though slightly lower results in the OLS regression. As a first-step in accounting for differences in propensity to patent across these firms which are drawn from all manufacturing sectors, we add a dummy variable for the scientific sector which includes firms in the drug, computer, scientific instruments, chemical, and electrical equipment industries, and a proxy for firm size, 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 as 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 $u_{it} = n_{it} - \lambda_{it}$ we define the standardized residual as u_{it} divided by its estimated standard deviation: $\epsilon_{it} = (n_{it} - \lambda_{ig})/\sqrt{\lambda_{it}}$. We use these residuals to test our model specification in two ways.[4] First, the independence assumption can be tested by forming the 7×7 covariance matrix $\Sigma = (1/N)\sum_{i=1}^{N}(\epsilon_i \epsilon_i')$ where ϵ_i is a vector of residuals for firm i. The estimated correlation matrix has off diagonal element which equal .8 approximately. Significant correlation exists which casts serious doubts on the adequacy of our Poisson specification. 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 $\sigma_i^2 = (1/(T-1))\sum_t(u_{it} - \bar{u}_i)^2$ for $\bar{u}_i = (1/T)\sum_t \epsilon_{it}$ against $\bar{\lambda}_i = (1/T)\sum_t \lambda_{it}$. 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 $\log \sigma_i^2$ on $\log \bar{\lambda}_i$ takes the form, $\log \sigma_i^2 = -.68 + 1.42 \log \bar{\lambda}_i$. Thus, we need also to attend to this failure of our initial specification.

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." We know from the analysis of linear panel data models that there are two methods which can be used for this type of problem: random effects and fixed effects. We explore first the random effects specification. In the regression model this implies an equicorrelated

---

[4] One potential problem arises here. Since a common β is used to form u, under the null hypothesis of zero covariance of the true u_{it}'s, induced covariance of order $(1/NT)$ exists among the u_{it}'s. But since $NT = 896$ in our sample, this problem and the associated Cox-Snell [11] corrections are quite small.
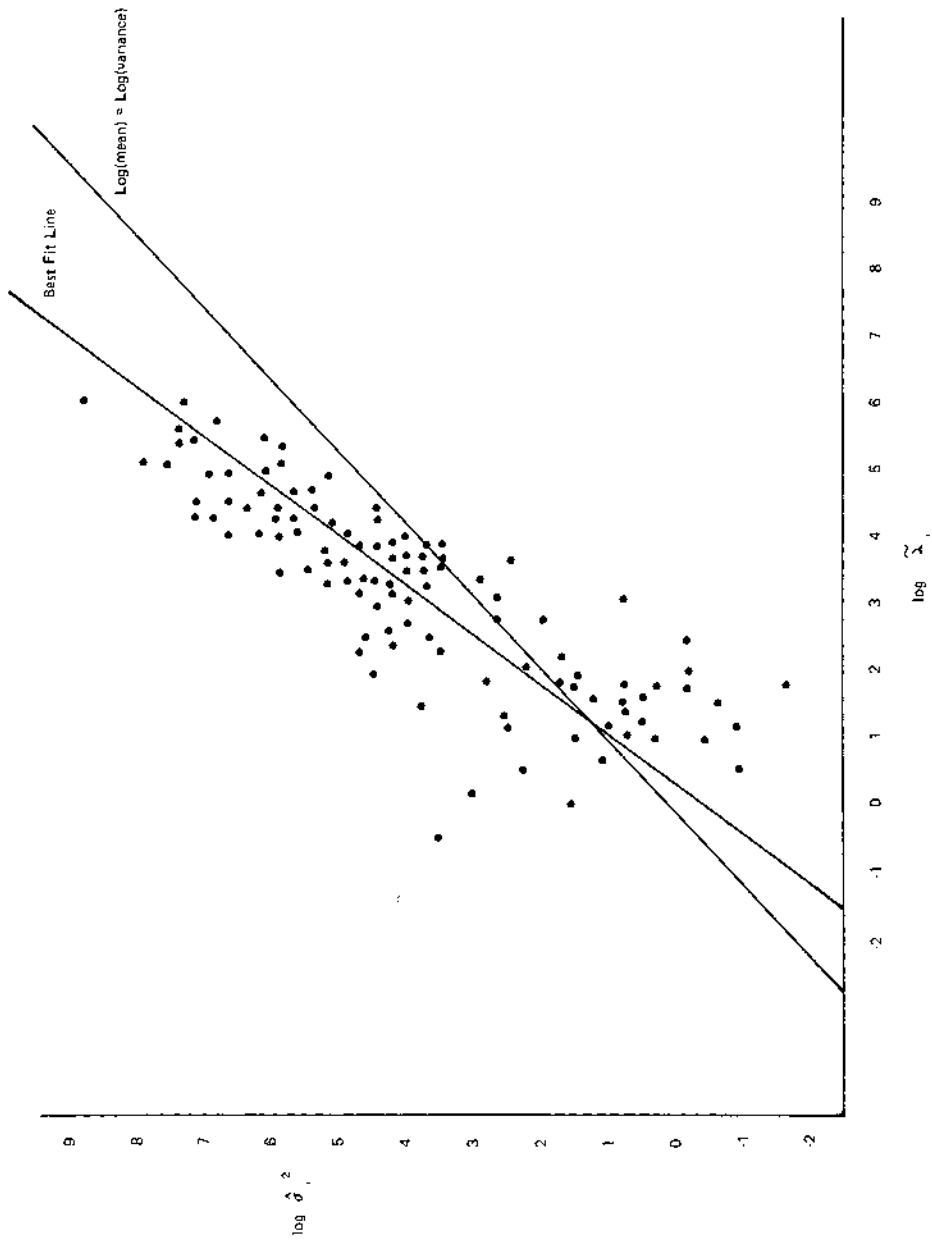
FIGURE 1—Within firm residual variance versus average lambda.

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 Anderson [1, 2]. 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, in the sense of diFinetti. 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 expenditures. The random effects specification is then no longer valid.[5] We use Hausman's [17] test to decide whether there exists a significant nonrandom correlation between the X_{it} and the $\tilde{\alpha}_i$'s.

We first consider the random effects specification. Because $\tilde{\lambda}_{it}$ needs to be positive, we write it in the form

$$\tilde{\lambda}_{it} = \lambda_{it}\alpha_i = e^{X_{it}\beta + \mu_0 + \mu_i}$$

where μ_i is the firm specific effect and μ_0 is the overall intercept. We include μ_0 in λ_{it} so that $Ee^{\mu_i} = 1$. The Poisson probability specification then becomes

$$(2.1) \qquad \operatorname{pr}(n_{it} \mid X_{it}, \mu_i) = \frac{e^{-\lambda_{it}e^{\mu_i}}(\lambda_{it}e^{\mu_i})^{n_{it}}}{n_{it}!}.$$

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

$$(2.2) \qquad \operatorname{pr}(n_{i1}, \ldots, n_{iT}, \mu_i \mid X_{i1}, \ldots, X_{iT})$$

$$= \operatorname{pr}(n_{i1}, \ldots, n_{iT} \mid X_{i1}, \ldots, X_{iT}, \mu_i) g(\mu_i)$$

$$= \prod_t \frac{\lambda_{it}^{n_{it}}}{n_{it}!} e^{-e^{\mu_i}\Sigma_t \lambda_{it}} e^{\mu_i \Sigma_t n_{it}} g(\mu_i),$$

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 variable with parameters (δ, δ), so that $E\alpha_i = 1$

[5] This problem has been recently discussed by Mundlak [21], Hausman [17], Chamberlain [7], and Hausman–Taylor [18]. Gourieroux et al. [13] emphasize problems which may arise if a particular distribution is chosen for the α_i.

and $V\alpha_i = 1/\delta$. As long as λ_{it} contains an intercept this normalization involves no loss of generality. We integrate by parts to find[6]

$$(2.3) \quad \mathrm{pr}(n_{i1}, \ldots, n_{iT} \mid X_{i1}, \ldots, X_{iT})$$

$$= \int_0^\infty \prod_t \left[\frac{\lambda_{it}^{n_{it}}}{n_{it}!} \right] e^{-\alpha_i \sum_t \lambda_{it}} \alpha_i^{\sum_t n_{it}} f(\alpha_i) \, d\alpha_i$$

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

where $\Gamma(\cdot)$ is the gamma function, $\Gamma(z) = t^{z-1} e^{-t}$ for $z > 0$. For this model the expectation of n_{it} is λ_{it} and the variance is $\lambda_{it}(\lambda_{it} + \delta)/\delta$. Therefore, the ratio of the variance to the mean is now $1 + \lambda_{it}/\delta$ so that the ratio grows with λ_{it}, which is what we observed in the residuals in Figure 1 although the observed relationship is not exactly quadratic as would be implied by the above formula. Maximum likelihood estimation of the parameters of the model in equation (2.3) is straightforward although we can no longer prove global concavity due to the addition of the δ parameter. 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 delta parameter using $V(\alpha) = 1/\delta$.

Results for the random effects Poisson specification are given in columns 1 through 3 of Table II. We see that the U-shaped lag structure of R&D is somewhat attenuated from that in Table I, but there is still a significant positive coefficient on the last lag. The total R&D effect is lower than that in our basic model in Table I although the exogenous time effect and the decline in the R&D coefficient over time remain about the same. The implied variance to mean ratio for patents at the means of the variables is about 20 and it grows with the estimated λ_{it}.

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 [24], Andersen [3], and Haberman [16]). Instead, we use the conditional maximum likelihood approach of Andersen [1, 2] and condition on the sum of patents $\sum_t n_{it}$. Since the Poisson distribution is a member of the exponential family, a

---

[6] Note that this specification is close to the classic Greenwood–Yule [14] specification which leads to a negative binomial specification. A similar probability specification was derived by Bates and Neyman [6] for a somewhat different model of accident proneness from that of Greenwood and Yule. Bates and Neyman named the distribution the multivariate negative binomial distribution. It is also referred to as the negative multinomial distribution.

TABLE II

ESTIMATES OF THE POISSON MODEL WITH FIRM EFFECTS

	Random Effects[a]				Fixed Effects[b]	
	1	2	3	4	5	6
$\log R_0$	.36 (.02)	.45 (.01)	.482 (.012)	.31 (.04)	.35 (.03)	.481 (.034)
$\log R_{-1}$	.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 (.0005)	.0368 (.0025)	-.03 (.003)	-.027 (.002)	.0369 (.0079)
time · $\log R_0$			-.0170 (.0006)			-.0171 (.0020)
log book value			.32 (.04)			
dummy (scientific sector)			.18 (.16)			
intercept	2.30 (.16)	2.65 (.011)	0.49 (.17)			
δ	1.20 (.15)	.98 (.12)	1.40 (.15)			
sum of $\log R$ coefficients	.59	.45	.414[c]	.43	.35	.413[c]
log likelihood	-3827.5	-3846.18	-3779.6	-3009.4	-3014.4	-2979.0
Tests for correlated firm effects				15.2	13.7	.01

[a] Random effects $\alpha_i = e^{\omega_i}$, α_i distributed independently as a gamma random variable with parameters (δ, δ).
[b] Fixed effects: Estimates conditional on the sum of patents over all T years.
[c] "Sum" evaluated at the midpoint of the period, 1971.

sufficient statistic exists for $T\tilde{\lambda}_i = \sum \tilde{\lambda}_{it}$ and it is $\sum_t n_{it}$. Since $\sum_t n_{it}$ is distributed as Poisson with parameter $\sum_t \tilde{\lambda}_{it} = \alpha_i \sum \lambda_{it}$, conditional maximum likelihood follows in a straightforward manner. Furthermore, it is known in the literature, e.g. Rao [26], that the distribution of n_{it} conditional on $\sum_t n_{it}$ gives a multinomial distribution

$$(2.4) \qquad \text{pr}\left(n_{i1}, \ldots, n_{iT} \mid \sum n_{it}\right)$$

$$= \text{pr}\left(n_{i1}, \ldots, n_{i,T-1}, \sum_{t=1}^{T} n_{it} - \sum_{t=1}^{T-1} n_{it}\right) \Big/ \text{pr}\left(\sum n_{it}\right)$$

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

Set $p_{it} = \tilde{\lambda}_{it}/(\sum_t \tilde{\lambda}_{it})$ and we have the multinomial distribution since $\sum p_{it} = 1$. Furthermore, for our particular specification we have

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

which is the so-called multinomial logit specification used by McFadden [20] in the discrete choice problem.[7] 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

$$(2.5) \qquad L(\beta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \Gamma(n_{it} + 1) - \sum_{i=1}^{N} \sum_{t=1}^{T} n_{it} \log \sum_{s=1}^{T} e^{-(X_{it} - X_{is})\beta}.$$

Equation (2.5) differs from the discrete choice likelihood function because here in general all the s_{it}'s are nonzero instead of only one nonzero value for the choice which is made. The gradient and Hessian for this likelihood are similar to those for the usual multinomial log specification: in particular, the Hessian can be shown to be globally concave by the Cauchy inequality provided the parameters are bounded, and hence its computations should and did converge rapidly.

The results for the conditional Poisson are given in columns 4 through 6 of Table II. The lag coefficients of R & D are now all small and insignificant with a $\chi_5^2 = 10$ for the test that they are jointly equal to zero. The U-shape of the

---

[7]Chamberlain [7] also derives a multinomial logit in his generalization of Cox's [10] fixed effects binomial logit model.

distributed lag no longer appears. The firm specific effect, α_i now represents both the accumulated stock of knowledge from past R & D in the firm[8] and unobserved permanent differences across the firms which affect their propensity to patent. Conditioning on permanent differences in the firms' levels of R & D expenditures has reduced the sum of the lag coefficients from .88 in the pooled model and .59 in the random effects model to .43.[9]

In Column 5 we estimate the model which contains only current R & D and find a coefficient of .35 which is 20 per cent below the sum of the coefficients in the previous specification. The time coefficient remains at −3 per cent 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 declining potency of R & D are repeated. Time itself has a positive coefficient of 4 per cent per year while the interaction with R & D has a coefficient of −.02.

To test whether our firm specific effects are correlated with the X in the model, we compare the random effects estimates to the fixed effects estimates using Hausman's [17] test. For the specification in columns 1 and 4, this test is 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 $\lambda_{it} = s_{it}\sum n_{it}$. Thus, under the null hypothesis of no serial correlation among the n_{it}, we have serial correlation of order $(-1/T)$ among the $\hat{u}_{it}$.

We form an asymptotic test as N becomes large by seeing whether the estimated covariance matrix from the residuals of the multinomial model of equations (2.4) and (2.5) has the form it would take under the null hypothesis of no serial correlation. Using the predicted probabilities

$$\hat{p}_{it} = e^{X_{it}\hat{\beta}} \Big/ \sum_{t=1}^{7} e^{X_{it}\hat{\beta}}$$

---

[8] 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.

[9] 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, between, within, 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 per cent for log Patents and 97 per cent 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.

we form the vector of each firm's predicted probabilities $\hat{p}_i = (\hat{p}_{i1}, \ldots, \hat{p}_{i7})$ and compute the multinomial covariance matrix:

$$(2.6) \qquad \tilde{\Omega}_i = \text{diag}(\hat{p}_i) - \hat{p}'_i \hat{p}_i \, .$$

Since $\tilde{\Omega}_i$ is singular by construction, we delete the first row and column to form a 6×6 matrix Ω_i. Likewise, we take the estimated residuals

$$(2.7) \qquad \hat{u}_{it} = \left(n_{it} - \hat{p}_{it} \left(\sum_{t=1}^{7} n_{it} \right) \right) \Big/ \left(\sum_{t=1}^{7} n_{it} \right)^{1/2} \qquad (t = 1, \ldots, 7)$$

and compute the covariance matrix $\tilde{S}_{i.} = \hat{u}_i \hat{u}'_i$. We delete the first row and column to form S_i. We then calculate the statistic

$$(2.8) \qquad R_i = \Omega_i^{-1/2} S_i \Omega_i^{-1/2'}$$

which should be close to the identity matrix if serial correlation is not present. To test for serial correlation we use the test statistic developed in Appendix B:

$$(2.9) \qquad Q = \frac{1}{N} \left(\sum_{i=1}^{N} m_i(\hat{\beta}) \right)' \hat{V}_m^{-1} \left(\sum_{i=1}^{N} m_i(\hat{\beta}) \right)$$

where $m_i(\beta)$ is the 15 element column vector composed of the unique nondiagonal elements of R_i in equation (2.8) and $\hat{V}_m$ is its asymptotic variance matrix which is calculated in Appendix B. The test statistic is computed to be 39.6. Under the null hypothesis this statistic is distributed as central χ^2_{15}. Having rejected the null hypothesis of independence we turn, in the next section, to the consideration of a more general model which allows for another source of within stochastic variation and which may be able to account for this apparent nonindependence.

3. NEGATIVE BINOMIAL MODELS

Even with the fixed effects Poisson model we still have the restriction that the variance and mean are equal, $En_{it} \doteq V(n_{it}) = \lambda_{it}$. On the other hand, the random effects Poisson has a variance to mean ratio of $1 + \lambda_{it}/\delta$ 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 Greenwood and Yule [14]. We then develop a fixed effects version of the negative binomial specification.

Greenwood and Yule 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 Greenwood and Yule. 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.[10] 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

$$(3.1) \qquad \text{pr}(n_{it}) = \int_0^\infty \frac{1}{n_{it}!} e^{-\lambda_{it}} \lambda_{it}^{n_{it}} f(\lambda_{it}) \, d\lambda_{it}$$

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

which is the negative binomial distribution with parameters (γ_{it}, δ). Computation of maximum likelihood estimates proceeds as for equation (2.3) 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 δ of about .05, implying a variance to mean ratio of 21 which is roughly consistent with the Poisson random effects model presented earlier.

Both estimates are higher than would be suggested by Figure 1 because the models we have used impose a constant variance to mean ratio across firms while the data suggest that the ratio grows with the number of patents. Another potential shortcoming of the negative binomial specification is that it does not allow for firm specific effects so that serial correlation of the residuals (i.e.,

[10]The parameter δ is different from its use in the last section.

nonindependence of the counts) may be a problem. We will return to this question later after we look at the results.

The estimates from the negative binomial specification are given in Table III. In the first two columns we consider specifications with and without lagged R & D. The total coefficient of R & D is about .75 for either specification, which is a decrease of 15 per cent from the corresponding model of Table I. We still find a large positive coefficient on the last lag of R & D. In column 3 we add the time and R & D interaction along with the scientific sector dummy variable and book value for the firm. Taking into account the time-R & D interaction, the estimated coefficient of current R & D falls from .55 in 1968 to .48 in 1974; the level is slightly lower and the decline slightly smaller than in the corresponding Poisson model. But as we suspected might happen, when we compute standardized residuals the problem of serial correlation reappears. 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 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 (2.4). 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. The derivation of the fixed effects negative binomial model is given in the Appendix. The resulting joint probability of a firm's patents conditional on

TABLE III
Estimates of the Negative Binomial Model

	Totals		
$\log R_0$	.43 (.08)	.75 (.02)	.56 (.03)
$\log R_{-1}$	− .04 (.14)		
$\log R_{-2}$	.16 (.14)		
$\log R_{-3}$	− .12 (.13)		
$\log R_{-4}$	− .07 (.15)		
$\log R_{-5}$	.41 (.10)		
time	− .05 (.01)	− .03 (.01)	.017 (.025)
time · $\log R_0$			− .012 (.006)
dummy (scientific sector)			.40 (.04)
log book value			.24 (.02)
intercept	− 1.10 (.07)	− 1.27 (.07)	− 2.20 (.13)
δ	.14 (.002)	.04 (.002)	.057 (.003)
sum of $\log R$ coefficients	.76	.75	.51[a]
log likelihood	− 3,820.8	− 3,845.3	− 3,747.4

[a] Sum evaluated at the midpoint of the period, 1971.

the seven year total is

$$(3.2) \qquad \mathrm{pr}\!\left(n_{i1}, \ldots, n_{iT} \mid \sum n_{it}\right)$$

$$= \left(\prod_t \frac{\Gamma(\gamma_{it} + n_{it})}{\Gamma(\gamma_{it})\Gamma(n_{it}) + 1}\right)\left[\frac{\Gamma\!\left(\sum_t \gamma_{it}\right)\Gamma\!\left(\sum_t n_{it} + 1\right)}{\Gamma\!\left(\sum_t \gamma_{it} + \sum_t n_{it}\right)}\right].$$

The log likelihood of the sample follows once we specify γ_{it}. We let the parameters of the underlying model be

$$(\gamma_{it}, \delta_i) = \left(e^{X_{it}\beta}, \phi_i/e^{\mu_i}\right)$$

where both ϕ_i and μ_i are allowed to vary across firms. The mean is

$$\tilde{\lambda}_{it} = \left(e^{X_{it}\beta + \mu_i}\right)/\phi_i$$

while the variance is

$$V(\tilde{\lambda}_{it}) = \left(e^{X_{it}\beta + 2\mu_i}\right)/\phi_i^2 .$$

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 corresponding unconditional negative binomial model we calculate

$$En_{it} = \left(e^{X_{it}\beta + \mu_i}\right)/\phi_i$$

with

$$V(n_{it}) = \left(e^{X_{it}\beta + \mu_i}/\phi_i\right)\!\left(1 + e^{\mu_i/\phi_i}\right)$$

so that the variance to mean ratio is $(e^{\mu_i} + \phi_i)/\phi_i$. 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.

Estimates for the fixed effects negative binomial model are given in the last three columns of Table IV. The coefficient of R & D is about one-half as large as the original negative binomial specification and is quite close to the conditional Poisson estimate. However, the standard errors on lagged R & D are much larger, reflecting the increased "noise" in the negative binomial specification. When we

TABLE IV

ESTIMATES OF THE NEGATIVE BINOMIAL MODEL WITH FIRM EFFECTS

	Random Effects			Fixed Effects		
log R_0	.45 (.06)	.52 (.025)	.49 (.04)	.41 (.06)	.37 (.03)	.39 (.04)
log R_{-1}	- .12 (.10)			- .15 (.10)		
log R_{-2}	.13 (.15)			.14 (.16)		
log R_{-3}	.01 (.19)			- .03 (.19)		
log R_{-4}	- .06 (.18)			- .07 (.17)		
log R_{-5}	.19 (.11)			.08 (.11)		
time	- .03 (.003)	- .02 (.003)	.01 (.01)	- .02 (.004)	- .02 (.003)	- .006 (.010)
time · log R_0			- .010 (.003)			- .004 (.003)
dummy (scientific sector)			.17 (.11)			
log book 71			.13 (.04)			
intercept	1.33 (.10)	1.47 (.10)	.64 (.22)			
a	2.56 (.36)	2.15 (.30)	2.56 (.35)			
b	1.72 (.34)	1.56 (.31)	1.74 (.37)	1.84 (.13)	1.86 (.11)	1.80 (.13)
sum of log R coefficients	.60	.52	.45[a]	.38	.37	.38[a]
log likelihood	- 3,303.1	- 3,310.6	- 3,304.9	- 2,467.3	- 2,468.9	- 2,468.5
Tests for correlated firm effects				580.	65.0	127.0

[a] Sum evaluated at the midpoint of the period, 1971.

interact time and R&D in column 6 we find important differences from the Poisson fixed effects model. First, the estimate of the coefficient of current R&D is .39, 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 since the negative binomial specification allows for an additional source of variance, the estimated standard errors are all larger, and the conclusions, while similar to the previous ones, are much less precise.

We again form a test for serial correlation. Define

$$\hat{f}_{it} = e^{X_{it}\hat{\beta}} \Big/ \sum_{t=1}^{7} e^{X_{it}\hat{\beta}}.$$

An extra term arises from the negative binomial (Dirichlet) assumption

$$(3.3) \qquad \hat{g}_i = \left(\sum_{t=1}^{7} n_{it} + \sum_{t=1}^{7} e^{X_{it}\beta} \right) \Big/ \left(1 + \sum_{t=1}^{7} e^{X_{it}\beta} \right).$$

Then we compute

$$(3.4) \qquad \hat{\Omega}_i = \hat{g}_i \big[\operatorname{diag}(\hat{f}_i) - \hat{f}_i'\hat{f}_i \big].$$

We again drop the first row and column to form Ω_i and use the residuals $\hat{u}_i$ to form $\tilde{S}_i$. The test statistic of equations (2.8) and (2.9) is used again. It is calculated to be 64.7 so that significant nonindependence is still present. It is interesting to note, however, that if we divide the sample on assets of $100 million, the test statistic equals 19.2 for the 44 small firms which is not significant for a χ^2_{15} random variable. But for the 84 larger firms the test statistic equals 58.5. The model is satisfactory for the small firms which created the "zero patent" problem, but trends in patents for a few quite large firms leave us with some serial correlation which is not explained by the model. While statistically significant, the serial correlation is not large with the r^2 between adjacent residuals of about .15 and its sign changing from positive to negative as the distance between observations increases.

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

$$(\gamma_{it}, \delta_i) = \left(e^{X_{it}\beta}, \phi_i/e^{\mu_i} \right)$$

so that both ϕ_i and μ_i vary across firms. Upon conditioning on the total number of patents in equation (3.2), 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/e^{μ_i} is randomly distributed across firms, independent of the X_{it}'s. An interesting difference exists between the Poisson random effects specifi-

cation and the negative binomial random effects specification. In the Poisson case, $\tilde{\lambda}_{it} = \lambda_{it}\tilde{a}_i$ where $\tilde{a}_i$ 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 $\epsilon_{it} = \alpha_i + \eta_{it}$.

We choose a distribution for $\delta_i = \phi_i/e^{t_h}$ which will allow us to integrate δ_i out of the marginal probability statement

(3.4) $\quad \mathrm{pr}(n_{i1}, \ldots, n_{iT} \mid X_{i1}, \ldots, X_{iT})$

$$= \mathrm{pr}(n_{i1}, \ldots, n_{iT} \mid X_{i1}, \ldots, X_{iT}, \delta_i)\, g(\delta_i)$$

where $g(\cdot)$ is the probability density of the incidental parameters. Because of the variance components, we need a two parameter distribution for δ_i and for ease of integration we take the ratio

$$\delta_i/(1 + \delta_i) = 1/(1 + e^{t_h}/\phi_i)$$

to be distributed as a beta random variable with parameters (a, b). Therefore, $\delta_i/(1 + \delta_i)$ has a density function

$$f(z) = \left[B(a,b) \right]^{-1} z^{a-1}(1 - z)^{b-1}$$

where $B(\cdot)$ is the beta function. The ratio $\delta_i/(1 + \delta_i)$ takes values on the unit interval which implies $\delta_i > 0$, which is appropriate for the scale parameter. The mean is

$$E(\delta_i/(1 + \delta_i)) = a/(a + b)$$

with variance

$$V(\delta_i/(1 + \delta_i)) = ab/(a + b + 1)(a + b).\,^{11}$$

We integrate using the beta density to find

(3.5) $\quad \mathrm{pr}(n_{i1}, \ldots, n_{iT} \mid X_{i1}, \ldots, X_{iT})$

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

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

---

[11] 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 equation (3.3).

where $z_i = \delta_i/(1 + \delta_i)$. Note that the last term in equation (3.5) corresponds exactly to a term in the fixed effects model of equation (3.3). But we now estimate additional parameters a and b from the beta distribution which describe the distribution of the δ_i across firms.[12]

Estimates of the random effects negative binomial specification are shown in columns 1 to 4 of Table IV. They fall in between the estimates from the totals model and the estimates from the fixed effects model. In the second column of Table IV, where only R&D and time are used in the specification of 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. With all five lags on R&D present, the estimates differ significantly from the fixed effects and totals estimates only in the last lag, which is where any firm effect due to presample R&D (truncation bias) will appear. The estimate of the time coefficients 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 variance to mean ratio of the effects is now estimated to be about 1.7 which is somewhat higher than the corresponding Poisson random effects estimate of about one. But now this ratio is being allowed to vary across firms rather than taking on a constant value as it does in the negative binomial totals or Poisson random effects models. A Hausman test of the random versus fixed effects specification yields 580 and 65 respectively for the first two specifications 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 the unexplained variance rises more than proportionately with predicted patents and hence with R&D.

In column 3 of Table IV 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 127.0, which clearly rejects the no correlation hypothesis, although the estimated coefficients are quite similar.

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 averages over time, for example $y_{i.} = (1/T)\sum y_{it}$. 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. But our conditional models differ from the

---

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

linear model because we no longer can use linear projections which separate the variables uniquely into X_i. and $X_{it} - X_i$. 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

$$(4.1) \quad \text{pr}(n_{i1}, \ldots, n_{iT} \mid X_{i1}, \ldots, X_{iT})$$

$$= \text{pr}\left(n_{i1}, \ldots, n_{iT} \mid X_{i1}, \ldots, X_{iT}, \sum_t n_{it}, \alpha_i\right)$$

$$\times \text{pr}\left(\sum_t n_{it}, \alpha_i \mid X_{i1}, \ldots, X_{iT}\right).$$

The first probability of the right hand side of equation (4.1) 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 (4.1). Therefore as we did in equation (2.3) we assume that $\alpha_i = e^{t_h}$ is distributed as a gamma random variable with parameters (δ, δ). We use the results of equation (2.3) on the sum of the patents $\sum_t n_{it}$ to derive the marginal probability

$$(4.2) \quad \text{pr}\left(\sum_t n_{it} \mid X_{i1}, \ldots, X_{iT}\right) = \left(\sum_t e^{X_{it}\beta}\right)^{\sum_t n_{it}} \left[\frac{\delta}{\sum_t e^{X_{it}\beta} + \delta}\right]^{\delta}$$

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

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 (4.2) instead of just X_i. appearing. Thus the between model does not depend on X_i. (or TX_i.) like the linear between model but instead depends on the within period variation via $\sum_t e^{X_{it}\beta}$, because of the nonlinearity introduced by the exponential functions. 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: $L(\beta, n_{i1}, \ldots, n_{iT})$ $= L_C(\beta, n_{i1}, \ldots, n_{iT} \mid \sum_t n_{it}) + L_M(\beta, \sum_t n_{it})$ 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 $L_C(\cdot)$ corresponds to the density in equation (2.4) and the marginal log likelihood L_M is the between model of equation (4.2). Similarly, the Fisher information regarding the parameters adds up, $J_T = J_C + J_M$ for $J_T = -\lim E(\partial^2 L/\partial\beta\partial\beta')$ 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 V 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 estimate of δ implies a variance to mean ratio for the seven year sum of patents of about 20 at the firm means, which is the same as the estimate we obtained for the random effects model on individual years of data. When we add the firm variables, however, this ratio is cut in half, in contrast to the results from individual years of data. 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.

For the negative binomial model the partitioning of the likelihood into conditional and marginal pieces is not as neat, however, since the form of the gamma functions allows us to identify the coefficients of variables which do not change over time from the conditional model. To see this, observe that if we include any variables which are constant over time in γ_{it} in equation (3.2), they will not necessarily cancel from the likelihood function and therefore their coefficients will be estimable. If we look at the Dirichlet derivation of the model, however, these variables do not really belong in γ_{it} since this derivation starts with the vector of patent shares in each year and treats them as random variables

TABLE V

ESTIMATES OF MARGINAL ("BETWEEN") FIRM MODELS

	Poisson		Negative Binomial	
$\log R_0$	.75 (.04)	1.18 (.15)	1.02 (.06)	1.29 (.15)
time $- \log R_0$		$-.26$ (.04)		$-.27$ (.04)
dummy (scientific sector)		0.20 (.18)		0.26 (.18)
log book value		0.29 (.10)		0.32 (.10)
intercept	1.72 (.08)	0.58 (.16)		
δ	1.29 (.15)	1.55 (.17)		
a			4.24 (.63)	2.82 (.49)
b			2.32 (.28)	3.17 (.69)
log likelihood	-806.5	-792.9	-802.4	-797.6

which are based on the underlying random variables λ_{it}, $t = 1, \ldots, T$. The λ_{it} may be freely rescaled by any factor which is constant over time leaving the shares unchanged. Therefore, the estimability of such coefficients is a kind of specification test of the model; failure of the test implies that the variance pattern in our data is not that implied by the Dirichlet or negative multivariate hypergeometric distribution.[13] A similar problem exists when we derive the between specification of the negative binomial model; because of the functional form of the gamma, we have identification of the coefficient of the time variable even though all we observe is the sum of patents over the seven years.

We now make the same assumptions on δ_i as we did for the derivation of the random effects negative binomial model of equation (3.3) to derive the between firm negative binomial model. 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,

(4.3) $\quad \mathrm{pr}\Big(\sum n_{it} \mid X_{i1}, \ldots, X_{iT}\Big)$

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

where a and b are the parameters of the underlying beta distribution. The log likelihood function for equation (4.3) follows directly. It is interesting to note that in equation (4.3) the leading terms in the numerator and denominator arise from the combinatorial term in the negative binomial distribution of equation (2.3) while the remaining terms arise from the ratio of two beta functions.

In columns 3 and 4 of Table V we give the results of the between negative binomial model of equation (4.3). By analogy to the between estimates for a variance components model, we can estimate only the overall variance of the model and not the decomposition into within firm and between firm variances. We accomplish this by dropping the intercept from γ_{it} and using the beta distribution to estimate the mean firm effect and its variance. The only difference between this model and the random effects Poisson model of columns 1 and 2 is the underlying distributional assumption on which each was based: the Poisson model variance arises only from the firm effect, whereas the negative binomial variance is a compounding of two effects which cannot be separated. The maxima of the likelihood functions for the two models are correspondingly close; in fact, for the second model the Poisson likelihood is higher. The coefficient estimates themselves are quite similar.

---

[13] We included the scientific sector and firm size variables in the model of column 6, Table IV and found that they were insignificant with a χ^2 of 2.2 with 2 degrees of freedom. The coefficients of interest (log R_0: .42 (.05), time: $-.005$ (.010), time log R_0: $-.004$ (.003)) do not change very much and we conclude that this form of misspecification is not a serious problem in our model.

5. SUMMARY

Our various models can be thought of as differing along two conceptual dimensions: (i) 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 (ii) are the relevant coefficients (β's) different when estimated in the conditional ("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 [21] language, are the individual "effects" *correlated* with the X's?

Table VI 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 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

TABLE VI
SUMMARY OF RESULTS

| Model | Log Likelihood | | Total R & D Coefficients[1] | |
	Poisson	Negative Binomial	Poisson	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	$-2,979.0$	$-2,468.5$	.41 (.03)	.37 (.04)
Sum of 2 and 3	$-9,044.2$	$-3,244.6$		
Tests of 2 and 3[1] against 1	$\chi^2_2 = 66.6$	$\chi^2_6 = 1006.$		
4. Totals (random effects)	$-3,779.6$	$-3,304.9$	.41 (.01)	.45 (.04)
5. Marginal (random effects)	-792.9	-797.6	.14 (.13)	.21 (.19)
Sum of 5 and 3	$-3,771.9$	$-3,266.1$		
Test of 5 and 3[1] against 4	$\chi^2_2 = 15.4$	$\chi^2_6 = 77.6$		

[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, $\beta_R + 4 \cdot \beta_{t \cdot R}$.

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 V 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. 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 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 [25] 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.

Another way of summarizing our results is to look at our estimates of the elasticity of patenting with respect to $R\&D$. They differ along two somewhat separate dimensions: (1) the implicit weighting of the individual observations—especially the random effects models versus the rest—with the former downweighting the larger observations (since they allow the variance to increase as the square of λ) and the differential treatment of zero values; and (2) what variables are included in the equation (none vs. 5 lagged $R\&D$ terms, size variables and sector dummies) and whether we allow for a correlation between firm effects (or past $R\&D$) and the included $R\&D$ terms. With the implied error variance proportional to λ (Poisson) we start with a total elasticity of about .9 which is reduced to .4 when all the various adjustments are made. It is still higher than the .3 OLS-within estimates because it makes a more proper allowance for the observed zero values. The difference between .4 and .9 can be decomposed roughly as follows: size and sector effects about .3; lagged $R\&D$ effects during the first five years about .07; effects of pre-sample $R\&D$ correlated within sample $R\&D$, about .08. In other words, while the current $R\&D$ component of the overall $R\&D$ elasticity of patents is .38, the overall sum is at least .55 (which is close to the Pakes and Griliches [25] estimates). It could be significantly higher, however, since we can only estimate that contribution of past $R\&D$ which is correlated with the included recent $R\&D$ terms.

The random effects model downweights the larger firms and starts out with a lower estimate of the total $R\&D$ coefficient (about .6), and reduces very much the influence of the size variables in the rest of the analysis. In this it is consistent with the results reported by Bound et al. [5] who showed for a larger cross-

sectional sample, that the estimated patents-R & D elasticity is quite sensitive to the weighting scheme used (or equivalently, that it is not really a constant elasticity relationship).

The rest of the conclusions are quite similar, however. The "pure" current R & D coefficient is around .38, time lagged R & D terms add another .08, while allowing for correlated effects of pre-sample R & D adds another .07 or so, yielding .53 as a lower bound on the total effect of R & D on subsequent patenting.[14] Of course, if one were willing to interpret observed size difference as the result of earlier R & D investments, then one could get an overall elasticity closer to unity (c.f. the marginal results reported in Table IV).

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. 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 so. 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.

Massachusetts Institute of Technology,
National Bureau of Economic Research, Palo Alto, California,
       and
Harvard University

APPENDIX A

To derive the fixed effects negative binomial model, we first find the moment generating function for the negative binomial distribution to be

$$m(t) = \left(\frac{1 + \delta - e^t}{\delta} \right)^{-\gamma}.$$

[14] These interpretations are based on using the observed evidence for truncation of the lag structure to attribute the correlation with firm effects to correlation with previous R & D whose effects decay very slowly.

Since the moment generating function of a 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

$$(A.1) \qquad \mathrm{pr}(w_1 \mid z = w_1 + w_2) = \frac{\mathrm{pr}(w_1)\mathrm{pr}(z - w_1)}{\mathrm{pr}(z)}$$

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

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

Note that in equation (A.1) 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

$$(A.2) \qquad \mathrm{pr}\left(n_{i1}, \ldots, n_{i7} \mid \sum n_{it}\right) = \left(\prod_t \frac{\Gamma(\gamma_{it} + n_{it})}{\Gamma(\gamma_{it})\Gamma(n_{it} + 1)}\right)\left[\frac{\Gamma\left(\sum_t \gamma_{it}\right)\Gamma\left(\sum_t n_{it} + 1\right)}{\Gamma\left(\sum_t \gamma_{it} + \sum_t n_{it}\right)}\right].$$

The marginal distribution of a given n_{it} conditional on $\sum n_{it}$ is a negative hypergeometric distribution (for integer values of the γ_{it}'s) so equation (A.2) is sometimes called a negative multivariate hypergeometric distribution for integer γ_{it}, e.g. Cheng Ping [8] and Johnson and Kotz [19].

We can also derive this distribution from the conditional Poisson model of the previous section, i.e., the multinomial distribution of equation (2.4). In that model the multinomial parameters

$$p_{it} = \lambda_{it}\Big/\sum_t \lambda_{it} = e^{X_{it}\beta}\Big/\sum_t e^{X_{it}\beta}$$

arose from the Poisson distribution. The natural mixing distribution for these 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 (2.4)

$$(A.3) \qquad \mathrm{pr}\left(n_{i1}, \ldots, n_{iT} \mid \sum n_{it}\right) = \frac{\left(\sum n_{it}\right)!}{\prod_t n_{it}!} E\left(\prod_t p_{it}^{n_{it}}\right)$$

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

where $\gamma_{i1}, \ldots, \gamma_{iT}$ are the parameters of the Dirichlet density. Note that equations (A.2) and (A.3) are identical as expected. The mean of equation (A.2) is $E(n_{it}/\sum_t n_{it}) = \gamma_{it}\sum_s n_{is}/\sum_s \gamma_{is}$, which is the same as $s_{it}\sum_s n_{is}$ 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,

$$V(n_{it}) = \left(\sum_t n_{it}\gamma_{it} \Big/ \sum \gamma_{it} \right)\left(1 - \gamma_{it} \Big/ \sum \gamma_{it} \right)\left(\sum n_{it} + \sum \gamma_{it} \right) \Big/ \left(1 + \sum \gamma_{it} \right).$$

We have again increased the variance over the multinomial case and made it grow with the expected number of patents to allow for overdispersion. The Dirichlet distribution occurs because each λ_{it} is distributed as a gamma random variable with parameters (γ_{it}, δ_i). It can be shown that the random variable $\lambda_{i1}/(\lambda_{i1} + \lambda_{i2})$ is distributed as a Beta random variable with parameters $(\gamma_{i1}, \gamma_{i2})$ for any δ_i which is the same across all t. The Dirichlet distribution is the multivariate generalization of the Beta distribution. 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 γ_{it}, $t = 1, \ldots, T$. Thus 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 p_{it}'s be random variables. Both derivations yield interesting insights into the basic model.

APPENDIX B: By JERRY HAUSMAN AND WHITNEY NEWEY[15]

To obtain a test for serial correlations in the fixed effects model, we use the fact that

$$E(u_{it}u_{is}) = \begin{cases} (1 - p_{it})p_{it}, & s = t, \\ -p_{it}p_{is}, & s \neq t, \end{cases}$$

where

$$p_{it} = e^{x_{it}\beta} \Big/ \sum_{t=1}^{7} e^{x_{it}\beta} \quad \text{and}$$

$$u_{it} = \left(n_{it} - p_{it}\left(\sum_{t=1}^{7} n_{it} \right) \right) \Big/ \left(\sum_{t=1}^{7} n_{it} \right)^{1/2}$$

Then for $p_i = (p_{i1}, \ldots, p_{i6})$ and $u_i = (u_{i1}, \ldots, u_{i6})$, we have

(B.1) $$E(u_i'u_i) = \text{diag}(p_i) - p_i'p_i = \Omega_i,$$

where the last equality defines Ω_i. We have deleted the last observation in forming p_i and u_i due to the fact that $\sum_{t=1}^{7} u_{it} = 0$. Equation (B.1) implies that

(B.2) $$E(\Omega_i^{-1/2}u_i'u_i\Omega_i^{-1/2}) = I_6$$

where I_6 is a six-dimensional identity matrix. A test for serial correlation can now be based on the sample counterpart of the off-diagonal elements of the 6×6 matrix $\Omega_i^{-1/2}u_i'u_i\Omega_i^{-1/2}$. Considering this matrix as a function β, let $m_i(\beta)$ be a 15 element column vector made up of the unique off-diagonal elements of this matrix. Then if the fixed effects model is correct, equation (B.3) will imply under the null hypothesis of no serial correlation that

(B.3) $$q = \left(\sum_{i=1}^{N} m_i(\hat{\beta})/\sqrt{N} \right)'[\hat{V}_m]^{-1}\left(\sum_{i=1}^{N} m_i(\hat{\beta})/\sqrt{N} \right) \xrightarrow{d} \chi^2(15)$$

as N gets larger for fixed T where $\hat{V}_m$ is a consistent estimator of the asymptotic covariance matrix of $\sum_{i=1}^{N}m_i(\hat{\beta})/\sqrt{N}$. Note that a first-order Taylor's expansion around the population value β_0 gives

(B.4) $$\sum_{i=1}^{N} m_i(\hat{\beta})/\sqrt{N} = \sum_{i=1}^{N} m_i(\beta_0)\sqrt{N} + \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\partial m_i}{\partial \beta}(\tilde{\beta}) \right]\sqrt{N}(\hat{\beta} - \beta_0),$$

[15]For a general treatment of this type of specification test, see Newey [23].

where $\tilde{\beta}$ lies between β_0 and $\hat{\beta}$. A central limit theorem and uniform convergence yields asymptotic normality of the statistic in equation (B.4) which leads to the χ^2 distribution of equation (B.3).

To estimate $\hat{V}_m$ let $l_i(\beta)$ the log of the likelihood for observation i, and $U_i(\beta) = \partial l_i(\beta)/\partial\beta$ be the score vector for observation i. The relationship of the test statistic and the score vector follows from

$$(\text{B.5}) \qquad E[\partial m_i(\beta_0)/\partial\beta] = -E[m_i(\beta_0)U_i(\beta_0)'].$$

Due to the presence of $\Omega_i^{-1/2}(\beta)$ in the definition of $m_i(\beta)$, $E[\partial m_i(\beta_0)/\partial\beta]$ will not be zero in general. However using equation (B.5) and an outer product estimator of the information matrix, a consistent estimation of V_m is

$$(\text{B.6}) \qquad \hat{V}_m = \frac{1}{N}\sum_{i=1}^{N} m_i(\hat{\beta})m_i(\hat{\beta})'$$

$$- \frac{1}{N}\sum_{i=1}^{N} m_i(\hat{\beta})U_i(\hat{\beta})'\left[\sum_{i=1}^{N} U_i(\hat{\beta})U_i(\hat{\beta})'\right]^{-1}\sum_{i=1}^{N} U_i(\hat{\beta})m_i(\hat{\beta})'.$$

It follows that the test statistic of equation (B.3) is analogous to a Lagrange multiplier test and can be computed via a regression.

REFERENCES

[1] ANDERSEN, E. B.: "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, B, 32(1970), 283–301.

[2] ————: "The Numerical Solution of a Set of Conditional Estimation Equations," *Journal of the Royal Statistical Society*, B, 34(1972), 42–54.

[3] ————: *Conditional Inference and Models for Measuring*. Copenhagen: Mentalhygiejnisk Forlag, 1973.

[4] ANDERSON, T.: *Multivariate Statistical Analysis*. New York, 1958.

[5] BOUND, J., C. CUMMINS, Z. GRILICHES, B. H. HALL, AND A. JAFFE: "Who Does R & D and Who Patents?" NBER Working Paper #908, Cambridge, 1982.

[6] BATES, G. E., AND J. NEYMAN: "Contributions to the Theory of Accident Proneness," *University of California Publications in Statistics*, 1. Berkeley: University of California Press, 1952, 215–253.

[7] CHAMBERLAIN, G.: "Analysis of Variance with Qualitative Data," *Review of Economic Studies*, 47(1980), 225–238.

[8] CHENG PING: "Minimax Estimates of Parameters of Distributions Belonging to the Exponential Family," *Acta Mathematica Sinica*, 5(1964), 277–299.

[9] CHERNOFF, H., AND E. LEHMANN: "The Use of the Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit," *Annals of Mathematical Statistics*, 25(1954), 579–586.

[10] COX, D. R.: *Analysis of Binary Data*. London: Methuen, 1970.

[11] COX, D. R., AND E. J. SNELL: "A General Definition of Residuals," *Journal of the Royal Statistical Society*, B, 30(1968), 248–275.

[12] GILBERT, G. C.: "Econometric Models for Discrete Economic Processes," paper given at the European Meetings of the Econometric Society, Athens, 1979.

[13] GOURIEROUX, C., A. MONFORT, AND A. TROGNON: "Pseudo Maximum Likelihood Methods: Application to Poisson Models," INSEE mimeo, Paris, 1981.

[14] GREENWOOD, M., AND G. U. YULE: "An Inquiry into the Nature of Frequency Distribution of Multiple Happenings," *Journal of the Royal Statistical Society*, A, 83(1920), 255–279.

[15] GRILICHES, Z., AND A. PAKES: "Estimation of Distributed Lags in Short Panels," NBER Working Paper #4, Cambridge, 1980.

[16] HABERMAN, S.: "Maximum Relationship Estimates in Experimental Response Models," *Annals of Statistics*, 5 (1977), 815–841.

[17] HAUSMAN, J. A.: "Specification Tests in Econometrics," *Econometrica*, 46 (1978), 1251–1272.

[18] HAUSMAN, J. A., AND W. E. TAYLOR: "Panel Data and Unobservable Individual Effects," *Econometrica*, 49(1981), 1377–1398.

[19] JOHNSON, NORMAN L., AND SAMUEL KOTZ: *Discrete Distributions*. New York: John Wiley, 1969.

[20] McFADDEN, D.: "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers of Econometrics*. New York: Academic Press, 1974.

[21] MUNDLAK, Y.: "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1978), 69–86.

[22] NELDER, J., AND R. WEDDERBURN: "Generalized Linear Models," *Journal of the Royal Statistical Society*, 135(1972), 370–384.

[23] NEWEY, W. K.: "Maximum Likelihood Specification Testing and Instrumented Score Tests," unpublished M.I.T. Ph.D. thesis, Cambridge, Mass., 1983.

[24] NEYMAN, J., AND E. L. SCOTT: "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1948), 1–32.

[25] PAKES, A., AND Z. GRILICHES: "Patents and R & D at the Firm Level: A First Look," *Economic Letters*, 5(1980), 377–381.

[26] RAO, C. R.: *Advanced Statistical Methods in Biometric Research*. New York: John Wiley, 1952.