NBER WORKING PAPER SERIES

# PATENTS AND R&D: SEARCHING FOR A LAG STRUCTURE


Bronwyn H. Hall

Zvi Griliches

Jerry A. Hausman



Working Paper No. 1227

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 1983

The research reported here is part of the NBER's research program in Productivity (World Economy). Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #1227
November 1983

# Patents and R&D: Searching for a Lag Structure

ABSTRACT

This paper extends earlier work on the R&D to patents relationship (Pakes-Griliches 1980, and Hausman, Hall, and Griliches, 1984) to a larger but shorter panel of firms. Using both non-linear least squares and Poisson type models to treat the problem of discreteness in the dependent variable the paper tries to discern the lag structure of this relationship in greater detail. Since the available time series are short, two different approaches are pursued in trying to solve the lag truncation problem: In the first the influence of the unseen past is assumed to decline geometrically; in the second, the unobserved past series are assumed to have followed a low order autoregression. Neither approach yields strong evidence of a long lag. The available sample, though numerically large, turns out not to be particularly informative on this question. It does reconfirm, however, a significant effect of R&D on patenting (with most of it occurring in the first year or two) and the presence of rather wide and semi-permanent differences among firms in their patenting policies.

Bronwyn H. Hall
National Bureau of Economic Research
204 Junipero Serra Boulevard
Stanford, California 94305
(415) 326-1927

Zvi Griliches
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, Massachusetts 02138
(617) 868-3900

Jerry A. Hausman
Department of Economics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
(617) 253-3644

August 1983

# PATENTS AND R&D: SEARCHING FOR A LAG STRUCTURE*

Bronwyn H. Hall, Zvi Griliches, and Jerry A. Hausman

## INTRODUCTION

In this paper we reexamine our earlier work on the relationship between R&D expenditures and patent applications (Pakes-Griliches 1980, and Hausman, Hall and Griliches 1984) using a larger sample and focusing primarily on trying to characterize the lag structure of this relationship. Earlier work had found a strong contemporaneous effect of R&D on patents but was inconclusive as to whether there was a significant lagged effect. Pakes and Griliches (1980), using the standard fixed effects model, found evidence of a lag truncation effect in the distributed lag of patents on R&D. That is, when they controlled for permanent differences across firms in the propensity to patent, the estimated coefficient on the last lag of R&D which they considered (R&D expenditures of four years prior) was significantly higher than the coefficients of more recent R&D. Hausman, Hall and Griliches (1984) used a different functional form (which took the discreteness of the patent data explicitly into account) and found similar results for the random (uncorrelated) effects model but not in their conditional fixed effects version. When they conditioned their estimates on the total number of patents received during the whole period, no coefficients except for the contemporaneous R&D variable were statistically significant either in the Poisson or negative binomial version.

Both studies used similar samples of about 120+ firms with seven to eight years of patent data and twelve to thirteen years of R&D data. In the meantime, a larger sample had become available (see Bound et al, 1982, and Cummins, Hall and Laderman 1983) and it was decided to try to investigate this question anew. Unfortunately, although this larger sample yielded consistent data for close to 750 firms, it is relatively short on R&D data. While we have patent data for the years 1967-77, the R&D data are available only back to 1972 for most of these firms and back to 1970 for about half of them. Thus, we cannot really tell whether there may be long delayed lag effects (longer than four years or so). What we will be looking for is whether there is a lag at all. The problem is complicated by our previous finding of persistent individual firm differences in their propensity to patent. The need to allow for such individual effects takes out much of the variance in the available short time series on R&D and makes it rather hard to distinguish between "firms differ because of their past R&D history" and "firms just differ" views of how these data were generated.

The other problem we have to deal with is the presence of a large number of zeroes in our dependent variable, the number of patents applied for in a particular year. We have dealt with this in two ways: (1) We use nonlinear least squares with robust standard errors on a model specified as $P_t = \exp(\Sigma p_\tau \log R_{t-\tau}) + \varepsilon_t$. This has the advantage of not requiring us to specify a distribution for $\varepsilon_t$, but we are unable to obtain conditional (fixed effect) estimates for this model due to its intrinsic nonlinearity and the shortness of our panel. Therefore, (2) we also chose to be explicit about the stochastic process generating $P_t$ and we estimated the comparable Poisson and negative binomial versions of this relationship, including the conditional versions of these models.[1]

The basic model that we will use in estimating and interpreting our various results is

$$(1) \qquad E(P_{it} \mid x_{it}, x_{it-1}, \ldots x_{it-\tau}, s_i) = e^{\beta' x + \delta' s + d_t + \alpha_i}$$

where the expected value of the number of patents applied for by firm i in year t (and ultimately granted) depends on the past history of its R&D expenditures ($x_{it} = \log R_{it}$), on permanent observable firm characteristics $s_i$ (such as size, measured by the net book value of its capital in 1972, and industrial specialization), unmeasured permanent differences in their propensities to patent $\alpha_i$, and changes in the overall average propensity to patent from year to year (across all firms) $d_t$. The error structure associated with different ways of estimating (1) will be discussed below. The main difference between various methods of estimation will be in the treatment of zeroes, the extent to which they recognize the intrinsic discreteness of the data, and the weight that they give to different observations. Since these issues were treated extensively in our previous papers, we shall allude to them only briefly as we go along. The main problem that we shall be dealing with in this paper is the lack of a long enough history on past R&D expenditures which will not allow us to estimate this model fully in an unconstrained fashion. To get somewhere we shall have to make strong assumptions either about the structure of the $\beta$'s as the lag increases or about the structure of the unseen past x's. Before we elaborate further on this, it will prove useful to rewrite (1) in greater detail and introduce the notion of the $\Pi$ matrix (Chamberlain 1980 and 1982) which summarizes the available (linear) information on the relationship between the dependent variable and all the available x's.

## II. THE MODELS

For ease of exposition, let us concentrate first on the (log) linear version of our model.

$$y_{it} = \beta_0 x_{it} + \beta_1 x_{it-1} + \dots + z_{it} + \alpha_i + e_{it} \tag{2}$$

where the $x$'s correspond now to the available information on past R&D expenditures while the $z_{it}$ represent the impact of the past (presample) unobserved history of the $x$'s as of time $t$. To simplify the exposition, we suppress the discussion of the other terms (firm characteristics and time dummies) in this section. $y_{it} = \log p_{it}$ is the dependent variable and $e_{it}$ is a random error (sampling or specification) distributed independently of the $x$'s. Both $z_{it}$ and $\alpha_i$ are unobserved and may be correlated with the included $x$'s. If they are, the $\beta$'s in (2) cannot be estimated consistently without making some additional assumptions and/or transformations. If there were no $z_{it}$, the $\beta$'s could be estimated consistently (barring other problems such as errors in variables) from the "within" part of the samples, from deviations around each firm's own means. We shall focus instead, at first, on the case where the $z_{it}$ are indeed important and the $\alpha_i$ are either absent or independent of the $x$'s. Even in this case (no or uncorrelated $\alpha_i$'s), we shall need to make strong assumptions about the $z_{it}$ to be able to identify the $\beta$'s. Two classes of assumptions can be used for this purpose, and we shall explore them both. The first assumes that the contribution of past history decays geometrically, at least after a few free terms in the lag structure. I.e.,

$$\begin{aligned} y_{it} &= \beta_0 x_{it} + \beta_1 [x_{it-1} + \lambda x_{it-2} + \lambda^2 x_{it-3} + \dots] + e_{it} \\ &= \beta_0 x_{it} + \beta_1 [x_{it-1} + \lambda x_{it-2} + \dots \lambda^{t-1} x_{i0}] + \lambda^{t-1} z_i + e_{it} \end{aligned} \tag{3}$$

where $x_{i0}$ is the first available $x$ (that is, we are assuming that we have available at least one lagged value of $x$) and $z_i$ is the impact of the unobserved past history of the $x$'s as of the initial first period, whose importance declines at the rate $(1-\lambda)$ over time. Now, if we had only three cross-sections we could rewrite (3) as

$$(3') \qquad \begin{aligned} y_{i3} &= \beta_0 x_{i3} + \beta_1 x_{i2} + \lambda\beta_1 x_{i1} + \lambda^2\beta_1 x_{i0} + \lambda^2 z_i + e_{i3} \\ y_{i2} &= \beta_0 x_{i2} + \beta_1 x_{i1} + \lambda\beta_1 x_{i0} + \lambda z_i + e_{i2} \\ y_{i1} &= \beta_0 x_{i1} + \beta_1 x_{i0} + z_i + e_{i1} \end{aligned}$$

Since $z_i$ is unobservable, estimating the observed part of these equations (separately for each cross-section or jointly) will result in biased coefficients, but the biases will be related and it may prove possible to unscramble them.

Instead of treating each of these cross-sections asymetrically, as implied by (3'), consider estimating a symmetric system, where each of the $y$'s depends on <u>all</u> the available $x$'s:

$$(4) \qquad \begin{aligned} y_{i3} &= \pi_{33} x_{i3} + \pi_{32} x_{i2} + \pi_{31} x_{i1} + \pi_{30} x_{i0} + v_{i3} \\ y_{i2} &= \pi_{23} x_{i3} + \pi_{22} x_{i2} + \pi_{21} x_{i1} + \pi_{20} x_{i0} + v_{i2} \\ y_{i1} &= \pi_{13} x_{i3} + \pi_{12} x_{i2} + \pi_{11} x_{i1} + \pi_{10} x_{i0} + v_{i0} \end{aligned}$$

The resulting coefficients, the $\pi$'s, summarize all the available information about the (linear) relationship between $y$'s and of the $x$'s in our sample. What is the expectation of these $\pi$'s ? To derive this, let us first define the projection of the excluded $z_i$ on all the available $x$'s:

(5) $$E^*(z \mid x_0,\ldots x_3) = \delta' x = \delta_3 x_3 + \delta_2 x_2 + \delta_1 x_1 + \delta_0 x_0$$

where $E^*$ is the wide sense expectation operator and the $\delta$'s are the projection or auxiliary equation coefficients, coefficients that summarize the relationship between the excluded $z$ and all the available $x$'s. Given (5), it is easy to see that the expectation of the estimated $\pi$'s is equal to

(6) $$E^*(\Pi) = \begin{bmatrix} \lambda^2\delta_3+\beta & \lambda^2\delta_2+\beta_1 & \lambda^2\delta_1+\lambda\beta_1 & \lambda^2\delta_0+\lambda^2\beta_1 \\ \lambda\delta_3 & \lambda\delta_2+\beta_0 & \lambda\delta_1+\beta_1 & \lambda\delta_0+\lambda\beta_1 \\ \delta_3 & \delta_2 & \delta_1+\beta_0 & \delta_0+\beta_1 \end{bmatrix}$$

and the twelve available $\pi$'s depend on only seven unknown parameters (2 $\beta$'s, 4 $\delta$'s, and $\lambda$). This matrix is in fact heavily constrained. For example,

$$\lambda = \pi_{23}/\pi_{13} = \pi_{20}/\pi_{10} = \pi_{30}/\pi_{20}$$

and it is obvious how one would recover the $\beta$'s. In practice, this set of equations is estimated jointly, imposing the non-linear constraints across equations. We shall use the non-linear analogs of the SUR procedure in TSP to estimate such models, allowing for arbitrary serial correlation of the $e_{it}$'s across time. This approach includes, as a special case, the "random effects," uncorrelated $\alpha_i$'s case.(2)

If we had just $\alpha_i$'s and no $z_i$'s (e.g. $\lambda = 0$), and the $\alpha$'s were correlated with the $x$'s, then by a similar argument

(7) $$E^*(\alpha_i \mid x_0,\ldots,x_3) = m'x$$

and the associated $\Pi$ matrix would be

$$(8) \qquad E^*(\Pi) = \begin{bmatrix} m_3+\beta_1 & m_2+\beta_1 & m_1 & m_0 \\ m_3 & m_2+\beta_0 & m_1+\beta_1 & m_0 \\ m_3 & m_2 & m_1+\beta_0 & m_0+\beta_1 \end{bmatrix}$$

This is the pure fixed (or correlated) effects case, which could also be estimated consistently from the "within" dimension of the data (in the linear case).

Assuming <u>both</u> the presence of $z$ ($\lambda \neq 0$) and correlated $\alpha$'s would produce a $\Pi$ matrix which would be a mixture of (8) and (6) and strain the identification potential of the data to its limit. With a larger number of cross-sections such a model might be estimable in principle but did not appear to be identifiable from the data available to us.

One should note two other possible estimation approaches to such a model. In the presence of the $\alpha_i$'s one may be tempted to first difference, with the result that (3') becomes

$$(9) \qquad \begin{aligned} dy_3 &= \beta_0 dx_3 + \beta_1 dx_2 + \lambda\beta_1 dx_1 + \lambda(\lambda-1)z + de_3 \\ dy_2 &= \beta_0 dx_2 + \beta_1 dx_1 + (\lambda-1)z + de_2 \end{aligned}$$

where $dx_t = x_t - x_{t-1}$ and the associated $\Pi$ matrix is

$$(10) \qquad E^*(\Pi) = \begin{bmatrix} \lambda c_3+\beta_0 & \lambda c_2+\beta_1 & \lambda c_1+\lambda\beta_1 \\ c_3 & c_2+\beta_0 & c_1+\beta_1 \end{bmatrix}$$

where the $c$'s are the coefficients in the projection of $(\lambda-1)z$ on the $dx$'s. While this model is identified ($\pi_{31}/\pi_{21} = \lambda$), the $c$'s are unlikely to be well defined (since there may be little correlation between the level variable $z$ and the subsequent $dx$'s) and an attempt to estimate all the coefficients jointly may experience severe convergence problems.

An alternative approach would take advantage of the geometric nature of the lag structure, and use lagged values of the dependent variable to solve out the unobserved $z_i$'s. Using the lagged dependent variables formulation would introduce both an errors-in-variables problem (since $y_{t-1}$ proxies for $z$ subject to the $e_{t-1}$ error) and a potential simultaneity problem due to their correlation with the $\alpha_i$'s (even if the $\alpha$'s are not correlated with the $x$'s). Instruments are available, however, in the form of past $y$'s and future $x$'s and thus such a system might be estimable along the lines outlined by Bhargava and Sargan (1983).

We do not pursue this line further here because we will be interested in estimating the nonlinear versions of our model, where neither the first difference or the lagged dependent variable option is available. In the nonlinear case, the first difference approach is equivalent to taking ratios, which founders on the presence of zeroes in our data. The lagged dependent variable approach is also not operational since the instrumental variable approach does not work for variables which are intrinsically nonlinear [i.e., there is no simple way to instrument the $(y_{t-1}-e_{t-1})^{\lambda}$ variable]. We shall return to this point further on, when we show that even though we cannot estimate $\lambda$ consistently, we can test the hypothesis $\lambda = 0$ using Lagrange Multiplier methods.

While the $\Pi$ matrix approach can be used also in the nonlinear context, it does require stronger assumptions to assure consistency. In particular, we shall have to assume the joint normality of the $z_i$ and $x_i$. That is, using (5) we can write

(11) $$z = \delta' x + \eta$$

and rewrite (1) as

$$(12) \qquad y_{it} = e^{(\beta+\delta)'x_i+\eta_i} + \varepsilon_{it} = [e^{(\beta+\delta)'x_i}]e^{\eta_i} + \varepsilon_{it}$$

For non-linear least squares procedures to be consistent under such circumstances we must assume independence between the $\eta_i$ and the x's. Non-correlation, which follows from the projection implicit in (11) is not enough.

The preceding used constraints placed on the lag distribution of the past history of the x's to achieve identification. Another approach to identification in such models is based on assumptions about the past history of the x's and does not require specific assumptions about the functional form of the lag distribution (see Pakes and Griliches 1982). Let us return to (3), free up the $\beta$'s, and forget the a's for a moment:

(13)

$$\begin{aligned} y_3 &= \beta_0x_3 + \beta_1x_2 + \beta_2x_1 + \beta_3x_0 \;\vdots\; + \beta_4x_{-1} + e_3 \\ y_2 &= \beta_0x_2 + \beta_1x_1 + \beta_2x_0 \;\vdots\; + \beta_3x_{-1} + \beta_4x_{-2} + e_2 \\ y_1 &= \beta_0x_1 + \beta_1x_0 \;\vdots\; + \beta_2x_{-1} + \beta_3x_{-2} + \beta_4x_{-3} + e_1 \end{aligned}$$

where for illustrative purposes, we have assumed a five term lag distribution with $x_{-1}$ through $x_{-3}$ constituting the relevant unobserved past history of the x's. The basic assumption that we shall make here is that the x's are generated by a relatively simple autoregressive (AR) process. If, for example, x's follow a first order AR, then in the projection of each of the unseen x's on all the available x's

$$(14) \qquad E^*(x_{-\tau}|x_0\ldots x_3) = g_\tau'x$$

only the coefficient of $x_0$ will be non-zero, since the partial correlation of $x_{-\tau}$ with $x_t$, <u>given</u> $x_0$, is zero for all $\tau > 0$ and $t > 0$. The $\Pi$ matrix for this case is thus

$$(15) \qquad E^*(\Pi) = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3+g_1\beta_4 \\ 0 & \beta_0 & \beta_1 & \beta_2+g_1\beta_3+g_2\beta_4 \\ 0 & 0 & \beta_0 & \beta_1+g_1\beta_2+g_3\beta_3+g_3\beta_4 \end{bmatrix}$$

and the first three $\beta$'s can be estimated consistently, leaving the last column of $\Pi$ free. If we had assumed that the $x$'s are AR(2), we would be able to identify only the first two $\beta$'s and would have to leave the last two columns of $\Pi$ free.(3)

Following Chamberlain, the basic procedure in this type of models is first to estimate the unconstrained version of the $\Pi$ matrix, derive its correct variance-covariance matrix allowing for the heteroscedasticity introduced by our having thrust the parts of the $\alpha_i$ or $z_i$ which are uncorrelated with the $x$'s into the random term (using the formulae in Chamberlain 1982, or White 1980), and then impose and test the constraints implied by the specific version deemed relevant.

Note that it is quite likely (in the context of longer T) that the test will reject all the constraints at conventional significance levels. This indicates that the underlying hypothesis of stability over time of the relevant coefficients may not really hold. Nevertheless, one may still use this framework to compare among several more constrained versions of the model to see whether the data indicate, for example, that "if you believe in a distributed lag model with fixed weights, then two terms are better than one."

## III. DATA AND RESULTS

The data we use are an extract from a larger and longer panel of firms in U.S. manufacturing drawn from the Compustat (Standard and Poor 1980). This dataset was assembled and combined with patent data from the Office of Technology Assessment and Forecasting at the NBER and is described in Bound et al (1984) and Cummins, Hall, and Laderman (1982). The original universe from which our sample comes consisted of approximately 2700 firms in the manufacturing sector in 1976, and included almost all of the firms which report R&D expenditures to the Bureau of Census-NSF R&D survey.

Our sample of firms was chosen from this universe by requiring that data on sales, gross capital, market value (value of common stock), and R&D be available for all years from 1972 through 1977 with no large jumps during that period. A jump is defined as an increase in capital stock or employment of more than 100 percent or a decrease of more than 50 percent. This test was not applied unless the change in employment was greater than 500 employees or the change in capital stock was greater than two million dollars. We also removed six firms which had abnormally small R&D values (less than $10,000) in one of the years. The number of firms remaining in the sample after these cuts was 738, with a size distribution heavily tilted toward the larger firms in our original universe. Table 1 shows the selectivity of this sample with respect to size and indicates that although we have only a quarter of our original sample of firms, most of those lost were either smaller or were not R&D-doing (and reporting) firms. Our coverage of the larger R&D firms is almost complete, and our sample includes 95 percent of the R&D dollars expended by the manufacturing sector in 1976.

Table 2 exhibits the characteristics of our remaining sample of firms, both the 738 firms with R&D between 1972 and 1977 and a subset of firms with a longer R&D history back to 1970. Quantiles are shown in order to give some indication of the skewness of the data: for example, median sales for this sample in 1976 was 177 million dollars, while mean sales was 979 million dollars. The subset of firms with a longer R&D history consists of somewhat larger firms and is more heavily tilted toward the scientific sector. Even for this sample of relatively R&D-intensive firms, we find that over 20 percent of the firms applied for zero patents in 1976 and that more than half applied for less than five. This confirms our impression that the patents variable in this data must be treated in a way which correctly reflects its relative imprecision at small values. Previous experience with estimation of the patents equation in the cross section (Bound et al 1982) has shown us that slope coefficient estimates may not be robust to changes in the way in which we specify the error in the equation (and the weighting which is implied by such specification).

Bound et al found that estimates of the elasticity of patenting with respect to R&D at the average R&D in the sample varied from .35 to 2, depending on the choice of specification: log linear, Poisson, negative binomial, or nonlinear least squares (exp(Xb)). This difference was greatly attenuated when the firms were divided into two groups, those with R&D budgets larger than two million dollars and those with smaller R&D budgets. In the present paper, the problem is not as severe, for two reasons: first, our sample is more heavily weighted toward the firms in the larger group (approximately 50 percent have R&D greater than two million, rather than 20 percent). Second, we have chosen to estimate a linear relationship between the log of patents and the log of R&D rather than the quadratic one in the previous paper. In addition, we present standard errors for the nonlinear least squares estimates which are computed without assuming anything about the disturbances except that they are additive and mean zero. That is, we assume that the model exp(Xb) is correct but we let the data tell us the form of the hetero-skedasticity. The formulas used are based on those of Eicker= White-Chamberlain.

Table 3 presents estimates of the nonlinear least squares model,

$$P_t = \exp\left(\sum_\tau \beta_\tau \log R_{t-\tau}\right) + e_t. \tag{16}$$

The estimates are obtained using the seemingly unrelated regression (SUR) method. Specifically, we estimate the covariance of the disturbances, $\hat{\Sigma}$ , using residuals computed from the unconstrained $\Pi$ matrix, and then use this estimate as weights when computing all the constrained models. This is a special case of generalized least squares, and the objective function is

$$\phi(\beta) = e(\beta)' \, (\hat{\Sigma} \times I_N)^{-1} e(\beta) \tag{17}$$

where $e$ is the "stacked" vector of residuals from the model of equation (16). This method allows for a free correlation over time for each firm, although it implicitly assumes that these correlation patterns are the same from firm to firm in estimating $\hat{\Sigma}$ . When we compute the standard errors for our estimates, we do not impose this assumption; each observation (firm) is weighted by its own residuals and their cross products, which allows for the possible heteroskedasticity across firms.

The first two rows of Table 3 present estimates of our most general model of section 2, given by equation (3). First we estimate a version with $\lambda$ equal to zero, which implies no $\beta$ coefficients after the first two, and then we give the version with $\lambda$ free. Note that this model is fit within the context of "random" rather than "correlated" firm effects, since the high degree of correlation in our $x$'s over time would make it difficult to discern both a decaying lag structure and an effect which has fixed coefficients with the $x$'s over time. The correlated effects question is addressed in the fourth row of the table where, in addition to contemporaneous R&D and R&D lagged once, we included all years of R&D with the same coefficients in each equation. The estimated lag coefficients do not change that much although the total effect drops from about .4 to .33, and a test of significance of the correlated effects yields an insignificant $\chi^2(6) = 4.4$, using robust standard errors. Accordingly, we feel reasonably confident that leaving out the correlated effects should not bias our results too much.

In fact, as can be seen in row 2, the model with geometric decay on the lag coefficients after the first two is preferred. The coefficient decays rather rapidly, 50 percent each year, but it is estimated with considerable imprecision, so this result also should not be taken too seriously. As we saw in section 2, another way to ask the same questions is to model the past history of the $x$'s, rather than of the coefficients. Although there is some evidence that an AR(1) process might do just as well, we choose to model them as an AR(2) process to be on the conservative side. Then the $\Pi$ matrix to be estimated has its last two rows free due to correlation of the first two $x$'s with the left out $x$'s. Otherwise, it has free lag coefficients on the diagonals and above, and zeroes below. The results for this version are given in row 3 of Table 3, and show not much evidence in the data of a lag longer than about two years.

There are several findings of interest in Table 3. First, the estimated total elasticity of patents with respect to R&D

expenditures is fairly stable across different versions, 0.33 to 0.43, except for the full model with geometric decay on the lag coefficients. In this model it rises to .51, suggesting that there is something, although it is small, in the firm's history of R&D expenditures that matters for patenting. Second, it appears that the effect of R&D peaks after one year. Even in the geometric lag case, the estimated average lag of patents applied for behind R&D expenditures is only one and one half years. Third, it is difficult to tell whether there is any significant lag beyond the first two years. In both the AR(2) case and the geometric lag model, the additional terms are at best only marginally significant.

To check on these conclusions, especially the last, we have redone the same computations for the half of our sample (N = 394) where we have data on R&D for two additional past years, 1970 and 1971. These results are shown in Table 4 and are inconclusive. The finding that correlated effects do not matter very much seems to hold up in these data (compare rows 1 and 4 again, and note that the goemetric decay model estimates almost the same total effect with a better fit). However, the total contribution of the lags beyond the first year in the $\lambda$ free model is small (about .13) and in the AR(2) version it is negative, albeit with large standard errors.

We turn next to the results of estimating the Poisson and negative binomial versions of our models. The advantage of these models is that they take explicitly into account the non-negativity and discreteness of our data. Moreover, the conditional versions of these models allow us to estimate a fixed effects model, something that we could not do easily with the nonlinear least squares estimates discussed in the previous section. On the other hand, because they are significantly more expensive to compute and more complicated to manipulate, we cannot really explore all the alternative hypotheses about the lag structure in their framework.

These models were described in detail in our earlier paper (Hausman, Hall, Griliches 1984) and we shall summarize only their main features here: The log likelihood function for the Poisson model is given by

$$\text{(18)} \qquad \log L = \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it}! - \exp(X_{it}\beta) + y_{it}X_{it}\beta]$$

and its conditional version is

$$\text{(19)} \qquad \log L = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \log \left[ \sum_{s=1}^{T} \exp[(X_{is}-X_{it})\beta] \right]$$

The Poisson estimates differ from the Nonlinear Least Squares ones reported in Tables 3 and 4 primarily by the weighting scheme used. The reported NLS estimates are unweighted, weighting implicitly the numerically larger deviations of the larger firms more than those of small firms. The Poisson estimates assume that the variance of the disturbances is proportional to the expected value of patents and weight the observations accordingly. The negative binomial version of the model generalizes the Poisson model by allowing for an additional source of variance above that due to pure sampling error. The logarithm of the likelihood for this model is

(20) $$\log L = \sum_i \sum_t \Big\{ \log \Gamma(e^{X_{it}\beta} + y_{it}) - \log \Gamma(e^{X_{it}\beta})$$

$$- \log \Gamma(y_{it}+1) + e^{X_{it}\beta} \log \delta - (e^{X_{it}\beta} + y_{it}) \log(1+\delta) \Big\}$$

where $\delta$ is the variance parameter ($Vy_{it} = e^{X_{it}\beta}/\delta$). The conditional version of this model conditions on the total number of patents applied for by the firm in all years:

(21) $$\log L = \sum_i \Big\{ \sum_t \log \Gamma(e^{X_{it}\beta} + y_{it}) - \log \Gamma(e^{X_{it}\beta}) - \log \Gamma(y_{it}+1) \Big\}$$

$$+ \log \Gamma(\sum_t e^{X_{it}\beta}) + \log \Gamma(\sum_t y_{it}+1) - \log \Gamma(\sum_t e^{X_{it}\beta} + \sum_t y_{it})$$

We estimate all these models using standard maximum likelihood techniques.

Since these models differ only by their distributional assumptions and not by the specification of the expected value they should all yield roughly the same results unless the basic specification of the model is wrong. In fact, it can be shown (see Gourieroux, Monfort, and Trognon 1981) that the NLS estimates are consistent even if the true distribution is Poisson and the Poisson estimates are consistent even if the true distribution is the negative binomial. Because they make different assumptions about the variance structure they do yield different estimates of standard errors, even in the case of similar coefficients.

Table 5 gives the major results of such computations. The first half of this table corresponds to the model estimated in our earlier paper and includes a time-R&D interaction in a search for possible changes in the "fecundity" of R&D over time. In general this interaction is not significant both because of the rather short period examined, the six years of 1972-77, and because there may not have been any systematic changes in the R&D coefficient over this period.[5] The second half of Table 5 corresponds to the models examined in Tables 3 and 4, with patents being a function of current and lagged R&D expenditures, but allowing also for permanent differences across firms in their patenting propensities. The results are rather similar except that in this format the first R&D coefficient is higher than the second and the estimated sum of the coefficients is somewhat lower in the Poisson case, although not in the negative binomial.

We turn now to the question whether there is any evidence for additional lags within this framework and we try to use the information contained in the lagged $y$'s, past patenting levels, to infer something about the importance of the unseen past. If the lag structure were geometric after the first two terms then we could solve out equation (3) for the missing $y$'s and substitute $y_{t-1}$ for them. However, what is needed here is the true "index value" of $y_{t-1}$, not its observed value which is subject to significant sampling error. While in the usual linear or log-linear models one could get around this by using instrumental variables, here, because of the intrinsic nonlinearity of $(y-e)^{\lambda}$ this does not really work. We turn, therefore, to a Lagrange Multiplier test of the hypothesis that $y_{t-1}$ belongs in the equation. The test itself is outlined in Appendix A. It is based on the computed residuals from the conditional Poisson model. These residuals, which are computed assuming that $y_{t-1}$ does not enter into the model, are then regressed on $\log y_{t-1}$. The coefficient in this regression should be zero if the null hypothesis is indeed correct. Since $y_{t-1}$ is subject to sampling error, the resulting regression coefficient may be attenuated and one may wish to use an instrumental variable estimation procedure here which is now consistent since the coefficient enters linearly in this equation. We use $\log y_{t-2}$ as an instrument, assuming that all the relevant serial correlation has been taken care of by the estimated fixed effects.[6]

The results of such computations are quite clear. For the equation reported in column 5 of Table 5, the estimated coefficient is .041 with an estimated t-ratio (using robust standard errors) of .16. Using an instrumental variable estimator the same numbers are .02 and .07 respectively. Thus, there is no evidence that there is any additional serial correlation or lagged x's effect left after one allows for permanent differences in the patenting propensity across firms.

An alternative way of asking this question is to look at the half of our sample where we have data on two more years of lagged R&D, back to 1970. This is shown in the last part of Table 5, where we see that including two more lagged logR terms in the conditional Poisson and negative binomial models neither improves the fit nor results in statistically significant coefficients. The conclusion remains the same: allowing for fixed effects it is not possible to estimate longer lag effects of R&D in these data. The significant effect that one can observe occurs in the first year or two.

There are at least two reasons for our failure to discern clear evidence of a longer lag structure from our data. First, the effects we are looking for are relatively small (relative to our ability to estimate them). Assume, for a moment, that the true total long run elasticity of patenting with respect to R&D expenditures is 1. We estimate that about .4 of it occurrs in the first two years and associate it with applied research and development expenditures. The effects of basic research take much longer, are more random, and hence are "smeared" over a longer period. Say that the rest of the effect, 0.6, is distributed over the next eight years. Then the average coefficient that we are looking for is .07, which is about the order of standard errors of such coefficients in our data. That is, the effect we are looking for is below the resolution power of our data.

The second related reason has to do with the properties of the R&D series in the real world. By and large, different firms have roughly constant (over a six year horizon) R&D budgets, which change from year to year, but largely randomly from the point of view of the sample as a whole. The first five serial correlation coefficients of log $R_{it}$ are estimated (using the MaCurdy (1982) approach) to be .99, .99, .96, .95 and .97 respectively, while the comparable serial correlations of the first differences are -.05, -.02, .01, .06; all not significantly different from zero. That is, we cannot reject the hypothesis that log $R_{it}$ follows a random walk. This should make it clear why we cannot estimate much of a lag structure without having a long history of data. There may be effects from the unseen past but we cannot learn about it from the observed present if it is largely uncorrelated with it.

## IV. CODA

We should not close this paper on the usual note of the failure of the data to live up to our econometric expertise. Even though we have not been able to elucidate the R&D to patents lag structure better, our overall findings are quite interesting, showing a persistent significant effect of R&D on patenting and rather wide and semi-permanent differences across firms in their patenting policies. The later finding provides the challenge for further and different style research: trying to understand how and why firms differ in their responses to the technological environment they find themselves in.

Table 1

| Sales | Number in 76 Cross Section All | Number in 76 Cross Section R&D>0 | Number in Sample | Coverage All | Coverage R&D>0 |
|---|---|---|---|---|---|
| less than $1M | 73 | 33 | 1 | .014 | .03 |
| $1M-10M | 548 | 293 | 21 | .038 | .07 |
| $10M-100M | 1102 | 579 | 261 | .24 | .45 |
| $100M-1B | 669 | 415 | 304 | .45 | .73 |
| $1B-10B | 204 | 167 | 141 | .69 | .84 |
| more than $10B | 12 | 11 | 10 | .83 | .91 |
| Total | 2608 | 1498 | 738 | .28 | .49 |

1976 R&D Expenditures
in 1976 dollars

| Sales | 76 Cross section | Sample | Coverage |
|---|---|---|---|
| less than $1M | 3.0 | 0.9 | .30 |
| $1M-10M | 65.3 | 5.3 | .08 |
| $10M-$100M | 525.2 | 266.3 | .51 |
| $100M-1B | 2354.1 | 2067.7 | .88 |
| $1B-$10B | 7830.6 | 7696.9 | .98 |
| more than $10B | 4593.2 | 4529.2 | .99 |
| Total | 15,371.3 | 14,566.3 | .95 |

Table 2

Key Variables in 1976

| Variable | Min | 1st Q | 738 Firms Median | 3rd Q | Max | 394 Firms Median |
|---|---|---|---|---|---|---|
| Sales ($M) | .6 | 57 | 177 | 674 | 49,000 | 238 |
| R&D ($M) | .02 | .69 | 2.2 | 9.7 | 1,256 | 3.5 |
| Patents | 0 | 1 | 3 | 14 | 798 | 4 |
| Fraction with zero patents | | | .22 | | | .21 |
| Fraction in scientific sector | | | .34 | | | .40 |

Notes to Table 2

All dollars are millions of 1976 dollars.

The scientific sector is defined as firms in the drug, computer, scientific instrument, chemical, and electric component industries.

Table 3: Patents and R&D: The Search for a Lag Structure

$P_{it} = \exp\,[\Sigma\beta_\tau \log R_{it-\tau}] + \varepsilon_{it}$ In 1 and 2 $\beta_\tau = \beta_1\lambda^{\tau-1}$, $\tau > 0$. N=738, T=6 (1972-77)

| | Estimated coefficients (Standard errors) | | | | Sum of Lag Coefficients | No. of Parameters | Trace Criterion |
|---|---|---|---|---|---|---|---|
| | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | | | |
| 1. $\lambda = 0$ | .179 | | | | .40 | 8 | 4889.0 |
| | (.077) | (.059) | | | | | |
| 2. $\lambda = .50$ | .111 | .200 | | | .51 | 15 | 4875/4 |
| (.18) | (.050) | (.074) | | | | | |
| 3. AR(2)x's | .121 | .194 | -.067 | .182 | .43 | 20 | 4746.2 |
| | (.084) | (.084) | (.075) | (.130) | | | |
| 4. Correlated effects, no lag truncation | .143 (.098) | .187 (.063) | | | .33 | 14 | 4858.0 |
| 5. Unconstrained $\Pi$ | | | | | | 36 | 4382.6 |

All equations contain year dummies, a dummy for scientific sector, and a size variable (Net plant book value in 1972). The coefficients of the 72 equation are free in all versions. In the $\lambda \neq 0$ versions, a set of eight additional parameters (six R&D terms, size, and sector) are constrained to decline as $\lambda^t$ across the cross sections (starting in 1973). In the correlated effects, no lag truncation, these parameters are constant over time. In the AR(2) version, the log $R_{72}$ and log $R_{73}$ coefficients are free in all years, while the size and sector dummies are fixed over time. Estimated using $\hat{\Sigma}$ from the unconstrained $\Pi$ version in the nonlinear SUR procedure in TSP. Robust standard errors computed using the Eicker-White-Chamberlain formulae.

Table 4: Patents and R&D: The Search for a Lag Structure

N = 394, T = 7 (1971-77)

| Equation | Estimated coefficients (Standard errors) $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | Sum of Log Coefficients | No. of Parameters | Trace criterion |
|---|---|---|---|---|---|---|---|---|---|
| 1. $\lambda = 0$ | .154 | .226 | | | | | .38 | 2 | 2797.0 |
| | (.091) | (.057) | | | | | | | |
| 2. $\lambda = .60$ | .150 | .085 | | | | | .36 | 11 | 2742.6 |
| (.19) | (.091) | (.053) | | | | | | | |
| 3. AR(2)x's | .060 | .285 | -.190 | -.021 | -.073 | -.196 | -.14 | 18 | 2671.7 |
| | (.090) | (.102) | (.070) | (.113) | (.063) | (.159) | | | |
| 4. Correlated effects, no lag truncation | .179 | .232 | | | | | .41 | 10 | 2757.2 |
| | (.125) | (.068) | | | | | | | |
| 5. Unconstrained $\Pi$ | | | | | | | | 48 | 2356.9 |

This table was estimated in the same way as Table 3, but on a subset of firms which reported R&D expenditures back to at least 1970. Here the parameters in the 72 equation are differentially constrained by the various models.

Table 5 : The Patents and R&D Relationship: Poisson and Negative Binomial Distribution Based Estimates

| Variables | N=738 | | | | | | N=394 | |
|---|---|---|---|---|---|---|---|---|
| | 1972-77 | | | | 1973-77 | | | |
| | Totals | | Fixed Effects | | Fixed Effects | | | |
| | Poisson | Negative B. | Poisson | Negative B. | Poisson | Negative B. | Poisson | Negative B. |
| log $R_t$ | .566 (.006) | .485 (.015) | .286 (.009) | .306 (.019) | .203 (.016) | .235 (.044) | .207 (.020) | .210 (.058) |
| log $R_{t-1}$ | | | | | .110 (.018) | .132 (.040) | .157 (.075) | .114 (.069) |
| log $R_{t-2}$ | | | | | | | -.011 (.025) | .041 (.063) |
| log $R_{t-3}$ | | | | | | | -.044 (.017) | .023 (.041) |
| $t \cdot \log R_t$ | .0012 (.0011) | -.0016 (.0024) | .0015 (.0014) | .0049 (.0012) | | | | |
| ℓnbkv72 | .210 (.003) | .147 (.012) | | | | | | |
| Sci Sect | .235 (.007) | .185 (.019) | | | | | | |
| Sum of Lag Coefficients | .57 | +.48 | .29 | .32 | .31 | .37 | .31 | .39 |
| Log L | 350,272 | 373,442 | -187,593 | -186,869 | -139,448 | -139,033 | -101,636 | -101,347 |

Notes to Table 5

The values in the table are the estimated coefficients of the respective variables and their standard errors (in parenthesis).

"Fixed Effects" -- Conditional estimates, conditional on the observed sum of patents for the period as a whole (for each firm). See Hausman, Hall and Griliches 1984, for more detail.

$\ell n$ bkv72 -- logarithm of the net book value of plant and equipment in 1972. A measure of size.

Sci. Sect. -- a dummy variable for "scientific sector" firms (consisting of firms in the drug, computer, scientific instrument, chemical, and electric component industries.)

## FOOTNOTES

*Prepared for the Conference on Quantitative Studies of R&D in Industry, Paris, September 9-10, 1983. We are indebted to the National Science Foundation (PRA 81-08635) and the National Bureau of Economic Research Program on Productivity and Technical Change for financial support. Elizabeth S. Laderman provided extremely able research assistance.

1. Note that nonlinear least squares is consistent even if $P_t$ is distributed as Poisson or Negative Binomial and the Poisson estimates are consistent even if the true distribution is Negative Binomial. See Gourieroux, C., A. Monfort and A. Trognon, 1981, and Hausman, Hall, and Griliches 1984.

2. The procedure we use is not fully efficient since we do not take into account in the estimation the possible heteroscedasticity across $i$ introduced by the projection of $z$ on $x$ and inclusion of the remainder in the new disturbance. (See Chamberlain 1982 for more detail.) We do, however, allow for such heteroscedasticity in computing our standard errors using formulae based on White 1980.

3. If the stochastic process generating the $x$'s can be assumed to be stable over time, efficiency could be improved by estimating the g's jointly with the $\beta$'s. See Pakes-Griliches 1982 for more discussion on this.

4. For comparison purposes with subsequent models, this one is estimated with the 1972 parameters left free and the $\delta$'s starting only from 1973 on.

5. The year constants do decline in 1976 and 1977 but this reflects the truncated property of our data. They are based on

total patents granted up to the end of 1979 and hence do not include some of the patents applied for in 1976 and 1977 to be granted after 1979. The year dummies decline by 7 and 21 percent respectively, relative to 1973-75, which is very close to the estimated decline in coverage: 4 and 21 percent respectively. (See Appendix Table 1.)

6. Here we define $y_t = p_t + .33$ to avoid the zeroes problem in $p_t$.

Appendix Table 1

The Distribution of Patents Applied for
by Date Granted: 1970 - 1977

| Year of Application | Years later 0 | 1 | 2 | 3 | 4 | 5+ | Total in Current Panel |
|---|---|---|---|---|---|---|---|
| 1969 | 0 | 11 | 66 | 20 | 2 | 1 | 100 |
| 1970 | 0 | 18 | 62 | 17 | 2 | 1 | 100 |
| 1971 | 0 | 18 | 64 | 16 | 1 | 1 | 100 |
| 1972 | 0 | 30 | 60 | 8 | 1 | 1 | 100 |
| 1973 | 1 | 43 | 47 | 7 | 1 | 1 | 100 |
| 1974 | 2 | 48 | 43 | 5 | 1 | 1 | 100 |
| 1975 | 2 | 49 | 41 | 6 | 1 | 1 | 99 |
| 1976[e] | 3 | 46 | 42 | 5 | * | * | 96 |
| 1977[e] | 1 | 41 | 37 | * | * | * | 79 |

Based on a sample of 100,000 patents from the 1969-79 OTAF tape on patents granted.

*Not computable

[e]Estimated

APPENDIX

In this appendix we develop a Lagrange Multiplier (LM) test for the presence of a lagged dependent variable in a Poisson type model. Since exponential models can always be written in generalized least squares form, we write the model which we wish to test as

$$\text{(A.1)} \qquad y_{it} = e^{X_{it}\beta+\alpha_i} z_{it-1}^{\lambda} + \varepsilon_{it}$$

where $z_{it-1}$ is the "true" value of the lagged dependent variable, equal to $y_{it-1} - \varepsilon_{it-1}$ . Because we do not observe $z$ and it enters nonlinearity into the equation, we cannot estimate (A.1) directly by instrumental variables. Instead we estimate equation (A.1) under the null hypothesis $\lambda = 0$.

$$\text{(A.2)} \qquad y_{it} = e^{X_{it}\beta+\alpha_i} + \varepsilon_{it}$$

and we do an LM test for $\lambda = 0$.

The gradient of the sum of squares function (likelihood function) for equation (A.1) with respect to $\lambda$ is

$$\text{(A.3)} \qquad \frac{\partial s}{\partial \lambda} = (e^{X_{it}\beta+\alpha_i} (\log z_{it-1}) z_{it-1}^{\lambda}) \varepsilon_{it}$$

Therefore, an approximate LM test is to take the estimated residuals from equation (A.2) and to do least squares with weights $e^{X_{it}\beta+\alpha_i}$ , on the equation:

(A.4) $$\hat{\varepsilon}_{it} = \Theta\ (\log y_{it-1}) + v_{it}.$$

The test is a significance test on the estimated coefficient $\hat{\Theta}$. We estimate equation (A.5) by instrumental variables to take account of the possible errors in variables problem, which arises from the fact that we use $\log y_{it-1}$ rather than $\log z_{it-1}$, $y_{it-1} = (z_{it-1} + \varepsilon_{t-1})$.

# REFERENCES


Bhargava, A. and D. Sargan. 1984. "Estimation of Dynamic Panel Models with Lagged Endogenus Variables," Econometrica, forthcoming.

Bound, J., C. Cummins, Z. Griliches, B.H. Hall, and A. Jaffe. 1982. "Who Does R&D and Who Patents?", National Bureau of Economic Research Working Paper #908, in R & D, Patents, and Productivity, 1984, Z. Griliches (ed.), Chicago: University of Chicago Press, forthcoming.

Chamberlain, Gary, 1982. "Multivariate Regression Models for Panel Data," Journal of Econometrics (18)1, 5-46.

________, 1978. "Omitted Variable Bias in Panel Data: Estimating the Returns to Schooling," Annales de l'Insee, #30-31, 49-82.

Cummins, C., B.H. Hall, and E. Laderman. 1983. "The R&D Master File: Documentation." Unpublished.

Eicker, F. 1967. "Limit Theorems for Regressions with Unequal and Dependent Errors, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Berkeley: University of California.

Gourieroux, C., A. Monfort and A. Trognon. 1984. "Pseudo-Maximum Likelihood Methods: Application to Poisson Models," Econometrica, forthcoming.

Hausman, J.A., B.H. Hall and Z. Griliches. 1984. "Econometric Models for Count Data With an Application to the Patents R&D Relationship," Econometrica, forthcoming.

MaCurdy, T. 1981. "Time Series Models Applied to Panel Data," Mimeo, Stanford University.

Pakes, A. and Z. Griliches. 1980. "Patents and R&D at the Firm Level: A First Look," NBER Working Paper #561, in R & D, Patents, and Productivity, 1984, Z. Griliches (ed.), Chicago: University of Chicago Press, forthcoming.

________. 1980. "Estimating Distributed Lags in Short Panels with an Application to the Specification of Depreciation Patterns and Capital Stock Constructs," NBER Working Paper No. 933.

White, Halbert, 1982. "Instrumental Variables Regression with Independent Observations," Econometrica 50, pp. 483-500.

______________, 1980. "Nonlinear Regression on Cross-Section Data," Econometrica 48, pp. 721-746.