INTERNATIONAL
ECONOMIC
REVIEW

June, 1986
Vol. 27, No. 2

PATENTS AND R AND D: IS THERE A LAG?*

BY BRONWYN H. HALL, ZVI GRILICHES AND JERRY A. HAUSMAN[1]

This paper analyzes the relationship between patenting and research and development activity at the firm level by the U. S. manufacturing sector during the 1970's. Previous work by Pakes and Griliches [1984a], which looked at a subset of the firms considered here, was the first attempt to use the patenting and R and D behavior of firms over time both to control for individual firm effects and to try to learn something about lags in the productivity of R and D. The present study extends their sample to 1979 and covers almost all of the firms doing appreciable amounts of R and D in the manufacturing sector. In attempting to characterize the lag structure of the patents-R and D relationship, a number of econometric problems arising from the panel nature of the data and from the measurement of the dependent variable have to be solved or at least considered in interpreting the results.

The basic model underlying this analysis has been described elsewhere [Pakes and Griliches (1984a), Hausman, Hall, and Griliches (1984)] and will only be sketched here. The annual research and development expenditures of a firm are considered to be investments which add to a firm's stock of knowledge. This stock of knowledge is depreciating over time so that the contribution of older R and D investment becomes less valuable as time passes. The aim of the study is to use patent applications in any given year as an indicator of the value of the additions to the underlying stock of knowledge, and to infer from the lag distribution on past R and D something about gestation lags in knowledge production. It is a maintained assumption of the model that patents are an indicator of the

* Manuscript received August, 1984; revised October, 1985.

[1] This is a revised version of NBER Working Paper #1227, prepared for the Conference on Quantitative Studies of R and D in Industry, Paris, September 9–10, 1983, held under the auspices of the Ecole Nationale de la Statistique et de l'Administration Economique and the National Bureau of Economic Research. We are grateful for helpful comments by the participants in that conference, especially by our discussants, Christian Gourieroux and Paul Stoneman, and by Mark Schankerman. We have also benefitted from the comments of two anonymous referees.
This research was supported by the National Science Foundation (PRA 81–08635) and the National Bureau of Economic Research Program on Productivity and Technical Change. Elizabeth S. Laderman provided extremely able research assistance.

output or "success" of R and D rather than simply the input of R and D. Testing this hypothesis requires another indicator of R and D success and is beyond the scope of this paper. The question has been investigated by Pakes [1985] and Griliches [1981] using the market value of the firm as an additional indicator with somewhat inconclusive results: in regressions of the rate of return of market value on the R and D history of the firm, contemporaneous patenting is moderately significant. This suggests that patents are measuring something above and beyond R and D inputs, which we identify as the "success" or output of R and D.

Patents are not the only output of R and D-they measure only a fraction of this output, and the fraction may vary considerably over industry and possibly also over time. Controlling for differences in the firms' propensity to patent (conditional estimation) as well as including an overall effect for each year are partial answers to this. A second and related problem with the existing data is that most of the information on the question will come from relative changes in the two variables over time within the firm; if these changes are contaminated by measurement error they are very small, the lag structure will be extremely difficult to discern.

A third problem has to do with the economic value of the patents themselves. Researchers such as Mansfield [1977] and Taylor and Silberston [1973] have suggested that the existence of the patent system may be a relatively unimportant factor in the research and development strategy of some firms. There is also a growing body of evidence [Grabowski (1983), Pakes (1984), Schankerman and Pakes (1984)] showing that a large fraction of patents granted are "worthless" or become worthless in a short period of time. This paper has very little to say on this range of topics; we observe the fact that firms do take out patents which are related to the output of their research and development laboratories (and other activities of the firms in the R and D area) and that therefore, patents can be used as an indicator of this activity in the aggregate even though the information conveyed by an individual patent may be very small.

In discussing an earlier version of this paper, Stoneman [1983] argued strongly that patents are an input to the R and D process rather than an output. That is, the patent application occurs at an early point in the development process and most of the expenditures that would be associated with it occur after the application is made. With these data we are in a position to investigate this question of timing, and find that the evidence for it is relatively weak, at least in aggregate firm behavior. The strongest thing one can say is that R and D and patents appear to be dominated by a contemporaneous relationship, rather than leads or lags.

The earlier work in this area also found a strong contemporaneous effect of R and D on patents but was inconclusive as to whether there was a significant lagged effect. Pakes and Griliches, using the standard fixed effects model, found evidence of a lag truncation effect in the distributed lag of patents on R and 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 and D which they considered (R and D expenditures of four years prior) was significantly higher than the coefficients of more recent R and D. Hausman, Hall, and Griliches 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 and D variable were statistically significant either in the Poisson or negative binomial version.

These previous studies both used samples of about 120 firms with seven to eight years of patent data and twelve to thirteen years of R and D data. For the current study, although we have fourteen years of patent data from 1966 to 1979, we have only eight years of R and D data (for 1972 through 1979) for about 650 firms, with an additional two years (1970 and 1971) for half of the firms. This fact constrains our ability to look for very long lag effects, especially since we cannot distinguish easily between permanent differences across firms in the propensity to patent and effects due to the unobserved past R and D history. We discuss this issue at greater length in the body of the paper.

The other problem we have to deal with is the specification of the error term in our model. The difficulty arises from two somewhat related causes: the presence of a large number of zeroes in our dependent variable, the number of patents applied for by a firm in a particular year, and the large size range of the firms in our sample. A previous paper [Bound *et al* 1984] which analyzed a large cross section sample, including the firms under consideration here, demonstrated that estimates were quite sensitive to the specification of the distribution of the error term. Since most of the estimators used were consistent, this can be construed as informal evidence of misspecification of the underlying model, possibly due to nonlinearity in the relationship of log patents to log R and D or to heteroskedasticity which is size-related. We have taken two approaches in obtaining our estimates in this paper: the first uses a nonlinear least squares specification with additive errors; for these estimates we are able to obtain robust standard errors which are correct in the presence of arbitrary heteroskedasticity, including year-to-year correlation within firms. For this version of the model however, we are unable to obtain conditional estimates, due to its intrinsic nonlinearity and the shortness of our panel. The second approach uses an explicit stochastic specification for the patents variable, that it follows a Poisson or negative binomial distribution, which enables us to obtain conditional estimates of the slope parameters, but at the price of a distributional assumption which may not hold.

The plan of this paper is the following: first we discuss the derivation of our dataset and look at the properties of our independent variable, R and D expenditures. Then we present some estimates of the basic patents-R and D relationship, followed by a discussion of the biases which may be present in the cross section estimation of this relationship. Finally we present conditional estimates of our model in an attempt to control for some of these biases and we conclude with a brief discussion of what we can learn from these data.

1. DATA

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 are required to report R and D expenditures to the Bureau of Census-NSF R and D survey. A few such firms which are privately held are excluded.

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 and D be available for all years from 1972 through 1979 with no large jumps during that period. A jump is defined as an increase in capital stock or employment of more than 100 per cent or a decrease of more than 50 per cent. 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 and D values (less than $10,000) in one

TABLE 1
SELECTION OF THE SAMPLE OF FIRMS

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	17	.031	.06
$10M–100M	1102	579	224	.20	.39
$100M–1B	669	415	259	.39	.62
$1B–10B	204	167	131	.64	.78
more than $10B	12	11	10	.83	.91
Total	2608	1498	642	.25	.43

1976 R & D EXPENDITURES IN 1976 DOLLARS

	76 Cross section	Sample	Coverage
less than $1M	3.0	0.9	.30
$1M–10M	65.3	4.7	.07
$10M–$100M	525.2	243.1	.46
$100M–1B	2354.1	1790.7	.76
$1B–$10B	7830.6	7224.1	.92
more than $10B	4593.2	4529.2	.99
Total	15,371.3	13,793.0	.90

of the years. The number of firms remaining in the sample after these cuts was 642, 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 and D doing (and reporting) firms. Our coverage of the larger R and D firms is almost complete, and our sample includes 90 per cent of the R and D dollars expended by the manufacturing sector in 1976.

Table 2 exhibits the characteristics of our remaining sample of firms, both the 642 firms with R and D between 72 and 79 and a subset of firms with a longer R and 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 were 182 million dollars, while mean sales were 1.06 billion dollars. The subset of firms with a longer R and D history consists of somewhat larger firms and is more heavily tilted toward the scientific sector. Even for this sample of relatively R and D-intensive firms, we find that over 20 per cent of the firms did not apply for patents in 1976 and that more than half applied for less than five. This confirms our impression that the patents variable in these 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* 1984] 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).

TABLE 2
KEY VARIABLES IN 1976

Variable	Min	1st Q	642 Firms Median	3rd Q	Max	346 Firms Median
Sales ($M)	.6	.57	182	760	49,000	263
R and D ($M)	.02	.73	2.3	11.0	1,256	3.8
Patents	0	1	3	18	831	5
Fraction with zero patents			.21			.17
Fraction in scientific sector			.37			.42

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.

In the later sections of this paper we look at this question again in an effort both to draw some robust conclusions from the data and to understand the reasons for the unstable coefficients. However, first we take a closer look at the behavior of our independent variable over time, since the way it evolves has important implications for our ability to identify the true lag coefficients in the presence of lag truncation and firm effects.

2. THE TIME SERIES BEHAVIOR OF R AND D EXPENDITURES WITHIN THE FIRM

To study the stochastic process for R and D, we use a procedure due to MaCurdy [1983] for computing the sample autocorrelation and partial autocorrelation functions. This method treats each firm as an independent draw on a time series process, so that we have 642 observations on the same short time series (8 years in our data). Since the autocorrelations and partial autocorrelations can be estimated consistently for each year due to the large sample size, it is not necessary to impose covariance stationarity. These variances and covariances

TABLE 3
TIME SERIES ANALYSIS OF LOG R & D[1]

Lag	642 Firms: Autocorrelations	642 Firms: Partial Autocorrelations	F-test for Equality of the Autocovariances
0	1.0	—	1.54
1	.987 (.051)	.992 (.002)	1.81
2	.991 (.051)	.054 (.035)	0.76
3	.974 (.051)	−.009 (.034)	2.51
4	.964 (.051)	.017 (.034)	2.35
5	.960 (.051)	−.036 (.032)	1.22
6	.959 (.052)	.006 (.032)	0.92
7	.959 (.052)	.055 (.123)	—

AUTOREGRESSIVE ESTIMATES FOR 1975–1979[2]

Equation	(1)	(2)	(3)	(4)	(5)
Log R_{-1}	.995 (.003)	.923 (.040)	.923 (.039)	.915 (.040)	.917 (.040)
Log R_{-2}		.074 (.039)	.082 (.053)	.067 (.040)	.069 (.040)
Log R_{-3}			−.009 (.034)		
Log P_0^3				.028 (.009)	
Log P_{-1}				.002 (.011)	.015 (.009)
Log P_{-2}				−.012 (.009)	−.002 (.009)
Standard Error	.292	.291	.291	.290	.291

Notes:
1. R and D expenditures are in millions of 1972 dollars. The deflator is described in Cummins *et al* (1982).
2. All equations contain a separate intercept for each year.
3. We have added 1/3 to the patents variable before taking the logarithm due to the presence of some zeroes.

are estimated by regressing the observed second moments of the series (centered, i.e., with year means removed) on a constant. The standard errors computed by such a regression are consistent under arbitrary heteroskedasticity of the underlying process since they correctly estimate the empirical variance of the second moments.

Using these estimates, we can test for equality of the Pth order covariances over time (we have eight estimates of the variance, seven for the first order covariance, six for the second, and so forth). In our data the F-statistic for the 28 implied restrictions is 1.42, which is insignificant at the conventional five per cent level. Accordingly, we impose stationarity in order to compute the autocorrelation and partial autocorrelation functions.

The results are shown in Table 3: the autocorrelations are all above 0.95, and show a very small decrease at longer lags, while the partial autocorrelations are essentially zero after the second lag, with the second lag equal to .054 (.035). This is strong evidence for a low order AR process; in fact, it is difficult to reject a random walk, although there is a hint of a small positive coefficient on the second lag and a first lag coefficient of slightly less than one. In order to check this result, we compute the AR regression itself and display the results in the second part of Table 3. The standard errors shown are heteroskedastic-consistent estimates, although they are in fact almost the same as conventional estimates, which is evidence that the assumption of constant within-firm variance is not a bad one. The basic result is that the AR2 specification can be accepted at conventional significance levels, and that the process is very close to a random walk.[2] We are not suggesting that a random walk is an adequate behavioral model of R and D investment, but merely that it provides a good description of the properties of the data, which we use later in the paper to help us disentangle the effects of pre-sample R and D.[3]

[2] Although the permanent-transitory model suggested by a referee fits fairly well in a panel as short as this one (with a residual variance only slightly higher than for the AR model), the properties of the residual covariance matrix over time suggest a clear preference for the AR model. The residual autocorrelations for this model are (−.05, .02, −.02, .05, −.002, −.05) as compared with (.49, .08, −.31, −.54, −.64, −.65) for a model consisting of a separate mean of R and D expenditures for each firm, time dummies, and a random noise term:

$$R_{it} = R_i + \alpha_t + \varepsilon_{it}.$$

Also, when the data for these firms are extended through 1981, the margin in favor of the AR1 model widens: its unexplained variance is .08 versus .12 for the permanent-transitory scheme. The coefficient of lagged R and D is 0.994.

[3] It is possible to derive this random walk behavior from a simple behavioral model of the firm: Following Abel [1984], Hall [1985] shows that for a risk neutral firm which maximizes the expected present value of cash flows subject to an accumulation equation for technology which is additive in R and D, and for which the production of technology is Cobb-Douglas in R and D expenditures, the optimal R and D expenditure each year is

$$R_{it} = R_{i,t-1} + u_{it}$$

where u_{it} is a function of the changes in real input and output prices and R_{it} is in natural logarithms. To the extent that price changes are constant across all firms (are captured by the time dummies), or have an expectation of zero for each firm this model corresponds to the one which we estimated.

The last two columns of this table provides a partial answer to the question of whether patents can be viewed as an input to the R and D process in this data. We use a simple version of a Granger causality test; with two lags of R and D used to predict the current level of R and D, we include contemporaneous and lagged log patents in the regression to see if they help in predicting R and D in the presence of its past history. The coefficient on contemporaneous log patents is significant ($t = 3.2$), but lagged patents are of no help in predicting future R and D, even if we leave contemporaneous patents out of the equation (last column). We tentatively conclude that there may be simultaneous movements in patents and R and D, but there is little evidence that past success in patenting leads to an increase in a firm's future R and D program above and beyond that implied by its R and D history.[4] We should qualify this result by noting that there is a considerably lower signal to noise ratio in the patents variable than in the R and D variable, both because of the skewness in patent values mentioned earlier and because it is intrinsically an integer variable. This has been well documented by Pakes and Griliches [1984a] and Pakes [1984]. Since R and D is highly correlated over time, it will be difficult to discern the independent contribution of patents to the R and D program in the presence of this noise.

3. BASIC RESULTS

In earlier work with the 1976 cross section of these firms, Bound *et al* found that estimates of the elasticity of patenting with respect to R and D at the average R and D in the sample varied from 0.35 to 2, depending on the choice of specification: log linear, Poisson, negative binomial, or nonlinear least squares. This difference was greatly attenuated when the firms were divided into two groups, those with R and D budgets larger than two million dollars and those with smaller R and D budgets. In the present paper, the problem is not as severe since our sample is more heavily weighted toward the firms in the larger group (approximately 50 per cent have R and D greater than two million, rather than 20 per cent), but it still persists and affects our estimates of the lag distribution.

In Table 3, we look at the differences in estimates of our basic model which are implied by differing specifications of the error structure. The model is

$$E(p_{it}|R_{it}, R_{i,t-1}, \ldots, s_i, t) = \exp\left[\sum_{\tau} (\beta_\tau \log R_{i,t-\tau}) + \delta' s_i + \alpha_t\right] \quad (1)$$

where s_i are the observed firm characteristics (size, as measured by the log of gross plant in 1972, and a dummy for the scientific sector). We use the years 1975 to 1979 for our 642 firms so that we can include three lags on R and D, yielding a total of 3210 observations on the dependent variable, patents. The first column shows the nonlinear least squares estimates of the parameters, which are obtained

[4] Ideally we would like to perform this test also in the other direction using patents on lagged patents and R and D, but there are difficulties in performing a comparable test due to the discrete nature of the dependent variable already alluded to.

by assuming an additive and homoskedastic error in equation (1). These estimates are consistent for the underlying coefficients, provided the model is correctly specified. The standard errors shown are robust to heteroskedasticity of the disturbances; they are computed using the formulas due to Eicker-White-Chamberlain, and allow both for differing variances across firms and arbitrary serial correlation over time within firms.

The next two columns of Table 4 give the results of estimating the Poisson and negative binomial versions of our models. The advantage of these models is that they take into account the non-negativity and discreteness of our data. Moreover, in the next section of this paper we will see that the conditional versions of these models allow us to estimate a fixed effects model, something that we cannot do easily with the nonlinear least squares version of the model. On the other hand, these models require us to be explicit about the exact form of the distribution from which the disturbance is drawn, and may produce inconsistent estimates if the distribution is not correct [Gourieroux, Montfort, and Trognon 1984].

TABLE 4

ESTIMATES OF THE PATENT EQUATION

642 Firms for 1975–1979

Variable	(1) Nonlinear least squares	(2) Poisson	(3) Negative binomial	(4) GMT		
	All Firms	All Firms	All Firms	All Firms	Small (306)	Large (336)
Log R_0	.12 (.30)	.28 (.03)	.21 (.07)	.30 (.10)	.31 (.11)	.32 (.15)
Log R_{-1}	.07 (.21)	.03 (.04)	.07 (.10)	.04 (.08)	.11 (.13)	.01 (.10)
Log R_{-2}	−.08 (.15)	−.001 (.036)	.08 (.10)	.06 (.08)	.14 (.13)	.02 (.10)
Log R_{-3}	.28 (.24)	.28 (.03)	.16 (.07)	.25 (.11)	.11 (.11)	.31 (.14)
Sum log R	.39 (.09)	.58	.52	.66 (.05)	.66 (.08)	.66 (.08)
Log book plant in 1972	.23 (.07)	.21 (.004)	.14 (.013)	.19 (.04)	.13 (.07)	.16 (.12)
Dummy (sci. sector)	.36 (.23)	.30 (.01)	.28 (.03)	.21 (.11)	−.13 (.16)	.25 (.12)
δ			.051 (.001)			
Log likelihood		280,034.	297,016.			

Notes:

1. All equations have a separate intercept for each year.
2. Standard errors for NLS and GMT are "robust" estimates computed by generalized Eicker-White-Chamberlain formula.

The Poisson and negative binomial models were described in detail in our earlier paper [Hausman, Hall, and Griliches 1984] and we shall summarize only their main features here. The log likelihood function for the Poisson model is given by

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

where y_{it} is the observed number of patent applications for a firm in a year and the X_{it} are the independent variables, R and D and firm characteristics. Estimates obtained with this model differ from the nonlinear least squares estimates primarily by the weighting scheme used. The NLS estimates are unweighted, implicitly weighting the numerically larger deviations of the larger firms more than those of the small firms. The Poisson model assumes that the variance of the disturbances is proportional to the expected value of the patents and weights the observations accordingly. The negative binomial 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

(3) $$\log L = \sum_{i=1}^{N} \sum_{t=1}^{T} \log \Gamma(\lambda_{it} + y_{it}) - \log \Gamma(\lambda_{it}) - \log \Gamma(y_{it} + 1) + \lambda_{it} \log(\delta) - (\lambda_{it} + y_{it}) \log(1 + \delta)$$

where $\lambda_{it} = \exp(X_{it}\beta)$ and δ is the variance parameter $(Vy_{it} = \exp(X_{it}\beta)(1+\delta)/\delta)$. We estimate both of these models by standard maximum likelihood techniques.

Finally, in the fourth column of Table 4, we show estimates computed using the quasi-generalized pseudo maximum likelihood [QGPML] method of Gourieroux, Montfort and Trognon [1984, henceforth referred to as GMT]. This estimator is based on the following idea: suppressing the t subscript for the moment, we assume that the true model for patents is

$$y_i = \exp[X_i\beta + e_i]$$

where $\exp(e_i)$ is a multiplicative disturbance drawn from an unspecified distribution. If a constant term is included as one of the X's, we can assume $E[\exp(e_i)] = 1$ and $V[\exp(e_i)] = \eta^2$. Then the expected number of patent applications conditional on the X's is $\exp(X_i\beta)$ and the variance is $\exp(X_i\beta) + \eta^2 \exp(2X_i\beta)$. That is, the variance equals the mean plus a parameter times the mean squared. We can obtain consistent estimates of the parameters β for this model using nonlinear least squares, use these to estimate η^2, form a vector of GLS weights which are proportional to the variance of the dependent variable:

$$w_i = y_i + \eta^2 y_i^2$$

and use these weights to obtain more efficient estimates of the β's. The formula for the variance of these estimates is given in GMT, and is a special case of the Eicker-White-Chamberlain formula with known weights.

Since all of these models differ only by their distributional assumptions and not by specification of the expected value, they should all yield roughly the same results unless the basic specification of the equation is wrong. In fact, it can be shown [see GMT] that both the NLS estimates and the Poisson estimates of the parameters are consistent if we have correctly specified the expectation in (1) and the true conditional distribution satisfies certain regularity conditions given in their article. Because the estimators make different assumptions about the error structure they do yield different estimates of the standard errors, even in the case of similar coefficients. In this respect, the nonlinear least squares estimates, weighted or unweighted, are the most robust, since we have computed standard errors which allow for unknown heteroskedasticity. It can be seen from the table that in return for making a relatively mild assumption about the form of the variance (that it is increasing in the mean and mean squared), we obtain a considerable increase in the precision of our estimates (compare 1 and 4).

The results of using the four different estimators shown in Table 4 are qualitatively the same, although there is a substantial increase in the coefficient on contemporaneous R and D as we move from nonlinear least squares to weighted nonlinear least squares (GMT). Since the estimators in columns 1, 2, and 4 are consistent if we have the correct model, but are estimated with different weighting schemes,[5] one possible explanation of the differences in coefficients, particularly the sum, may be that the relationship between patents and R and D is not stable across the firms in our sample. An indication that this is a problem is provided by the substantial increase in standard errors when we use robust estimates.

The dimension along which the weighting schemes vary is basically related to the size of the firms in the sample. Therefore we partitioned them into roughly equal groups: those with assets (book value of net plant) less than 25 million dollars in 1972 and those with assets greater than 25 million. We then estimated the same model on the two groups separately; the results are shown in the last two columns of Table 4. Although the total R and D effect is the same for the two groups (.66), it is distributed differently across the lags. This suggests that the maintained hypothesis of a roughly constant lag structure across the firms may be one reason for the apparent instability of our results. Unfortunately it is not possible with

[5] Another way to understand what the different estimators are doing is to examine the first order conditions (again suppressing the t subscript and writing e_i for y_i-exp $(X_i\beta)$):

NLLS: $\sum_i \exp(X_i\beta) \quad e_iX_i=0$

Poisson: $\sum_i 1 \qquad e_iX_i=0$

Negative binomial: $\sum_i (1+\eta^2 \exp(X_i\beta))^{-1}e_iX_i=0$

GMT: $\sum_i \frac{\exp(X_i\beta)}{y_i+\eta^2y_i^2} \quad e_iX_i=0$

Note that the first order condition which we show for the negative binomial model is *conditional* on the choice of η^2. Since η^2 is being estimated simultaneously this is not the full set of first order conditions for the problem; we merely include it for illustrative purposes. Joint estimation of η^2 is precisely what makes this estimator inconsistent when the distribution is not negative binomial, although the other three estimators remain consistent in this case since they are all versions of weighted least squares.

Displaying the first order conditions in this way reveals that the estimators only differ in their choice of weights, although NLLS and GMT are minimum distance estimators, and Poisson and negative binomial are maximum likelihood estimators. They can be ranked by the weight which they give to firms with larger X's, this dataset to construct a more detailed behavioral model which is capable of accounting for different lag structures across firms. We can only suggest areas for future investigation.

4. CORRELATED EFFECTS OR LAG TRUNCATION BIAS?

In obtaining the results shown in Table 4, there was no attempt to control for permanent differences in the propensity to patent across firms, except for the firm size variable and a dummy for the scientific sector. We expect that these differences may bias our estimate of the R and D coefficients if they are correlated with the R and D variables. All of the estimates, except possibly those for the smaller firms, exhibited evidence of a *u*-shaped lag structure, with the first and last coefficients being larger than those in the middle. The large coefficient on the last lag could be due to the correlation of the last R and D variable with earlier left-out R and D, but it turns out that under reasonable assumptions on the R and D process itself, it could also be caused by (correlated) fixed effects.

Assume that the log deflated R and D variable itself follows a first order autoregressive process:

$$R_t = \gamma R_t + e_t, \qquad e_t \text{ a white noise process}$$

Then the autocorrelation coefficients for the R and D process are $(1, \gamma, \gamma^2, \ldots)$ and the partial autocorrelation coefficients are $(\gamma, 0, 0, \ldots)$. We have seen that just such a pattern is consistent with our R and D data. If we maintain the hypothesis that R and D follows an AR1 process, we can compute the bias formula for the coefficients on R and D in the presence of two types of omitted variable: 1) pre-sample R and D which is correlated with in-sample R and D, and 2) a permanent fixed effect which has the same correlation with R and D in all periods. In the first case, for example, if the most recent pre-sample R and D belongs in the equation in addition to the included R and D, the bias formula for $\beta = (\beta_0, \beta_1, \ldots, \beta_k)$, where there are k lags in the regression, is

$$\text{(4)} \qquad \operatorname{plim} \beta = \beta + \frac{\beta_{k+1}}{\sigma_x^2(1-\gamma^2)} \begin{bmatrix} 1 & -\gamma & 0 \cdots & 0 & 0 \\ -\gamma & 1+\gamma^2 & -\gamma \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \cdots 1 & +\gamma^2 & -\gamma \\ 0 & 0 & 0 \cdots & -\gamma & 1 \end{bmatrix} \begin{pmatrix} \gamma^k \sigma_x^2 \\ \gamma^{k-1} \sigma_x^2 \\ \vdots \\ \gamma \sigma_x^2 \end{pmatrix}$$

$$= \beta + \beta_{k+1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \gamma \end{pmatrix}$$

Therefore the estimate of the last coefficient will be biased upward by $\beta_{k+1}\gamma$ where β_{k+1} is the coefficient on the R and D one period before the sample begins. If more than one lag of pre-sample R and D belongs in the equation, the bias on the last coefficient will be equal to $\sum \beta_{k+\tau}\gamma^{\tau}$, where τ indexes the pre-sample R and D lags. In our case, since γ is close to unity, we expect the last coefficient to be roughly equal to the sum of the lag coefficients for all the earlier R and D plus its own coefficient. Even if the R and D process were autoregressive of order two, we would expect approximately the same result. In that case both the last and next to last coefficient would be biased upwards, but because of the large first order serial correlation of the R and D process, the bias on the last coefficient would be ten times that on the next to last.

On the other hand, if we assume a fixed effect μ_i for each firm has been omitted, we obtain the formula

$$(5)\qquad \operatorname{plim}\beta = \beta + \frac{1}{\sigma_x^2(1-\gamma^2)}\begin{bmatrix} 1 & -\gamma & 0\cdots & 0 & 0 \\ -\gamma & 1+\gamma^2 & -\gamma\cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0\cdots 1+\gamma^2 & & -\gamma \\ 0 & 0 & 0\cdots & -\gamma & 1 \end{bmatrix}\begin{pmatrix} \eta\sigma_x\sigma_\mu \\ \vdots \\ \vdots \\ \eta\sigma_x\sigma_\mu \end{pmatrix}$$

$$= \beta + \frac{\eta\sigma_\mu}{(1+\gamma)\sigma_x}\begin{pmatrix} 1 \\ 1-\gamma \\ \vdots \\ 1-\gamma \\ 1 \end{pmatrix}$$

where η is the correlation of the fixed effect with R and D and σ_μ^2 is the variance of the effect. The implication is that we would see a large positive bias in the first and last coefficient and a smaller one in the middle coefficients. If γ is close to one, as it appears to be in our data, the bias for coefficients $\beta_2,\ldots,\beta_{k-1}$ would be negligible.

To explore this idea further, we use a longer sample of firms which have R and D data available back to 1970. This leaves us with 346 firms, slightly more than half of our original sample, and somewhat more heavily weighted towards larger firms. For this sample, we obtain estimates for a model with five lags on R and D (shown in Table 5). For comparison, estimates of the original model (with three lags) on this new sample are also shown. It appears from these results that some of the effect we observed in the last lag was indeed due to truncation (note how the coefficient on R_{-3} in column 1 is spread between R_{-3}, R_{-4}, and R_{-5} in column 2); if we push the idea further by estimating with seven lags on the last three years of data (1977 through 1979), the loading on the last lag seems to have disappeared. However, we have also pushed the data beyond the point where it will yield meaningful results, since significant instability in the R and D coefficients for adjacent years is now evident. In order to impose some stability on the lag structure in this case, we estimated the same relationship with the lag coefficients

TABLE 5
ESTIMATES OF THE PATENT EQUATION
346 Firms

Time period	Equation[1] (1) 1975–1979	(2) 1975–1979	(3) 1977–1979	(4)[2] 1977–1979
Log R_0	.16 (.16)	.19 (.16)	.34 (.23)	.13 (.17)
Log R_{-1}	−.02 (.10)	−.07 (.10)	−.30 (.18)	.05 (.08)
Log R_{-2}	.07 (.10)	.07 (.10)	.09 (.19)	.02 (.08)
Log R_{-3}	.36 (.18)	.06 (.09)	.02 (.13)	.03 (.04)
Log R_{-4}		.16 (.08)	.24 (.17)	.05 (.12)
Log R_{-5}		.17 (.12)	−.01 (.16)	.08 (.17)
Log R_{-6}			.10 (.12)	.11 (.08)
Log R_{-7}			.10 (.12)	.13 (.18)
Sum log R	.57 (.07)	.59 (.07)	.57 (.07)	.60 (.23)
Log book plant in 1972	.22 (.05)	.20 (.06)	.22 (.06)	.19 (.06)
Dummy (sci. sector)	.30 (.13)	.30 (.13)	.30 (.13)	.28 (.14)

Notes:
1. The estimation method is GMT, with separate intercepts for each year, and robust standard errors.
2. These coefficient estimates were obtained under the restriction that they lie on a cubic in the lag index.

constrained to lie on a cubic polynomial. These results are considerably smoother and are shown in column (4) of the table. Comparison of this column with columns 1 and 2 of the table shows that the lag distribution tends to drift backwards as we add lags, again suggesting the presence of lag truncation.

To examine the other alternative, a fixed effects explanation of the u-shaped lag distribution, we hypothesize a differing propensity to patent for each firm which is (possibly) correlated with its R and D activity. The reasoning in this section suggests that estimates conditional on the permanent patenting propensity of the firm should reduce both the first and last lag coefficient if we have correlated effects and only the last one if the problem is lag truncation (and the lag truncation is relatively constant from year to year). This leads us to look at models which are conditional on the permanent R and D behavior of firms in the next section.

5. CONDITIONAL ESTIMATES

We take two different approaches to obtaining conditional estimates for our model: the first, following Chamberlain [1984], includes all observed values of R and D (for a firm) in each equation with the individual lag coefficients constrained to be equal across the different years. This is an attempt to control for fixed effects which may be correlated with our R and D variables, since we cannot simply estimate the effects due to the shortness of our panel and the nonlinearity of the model.[6] The second approach imposes a specific distribution on the error term, namely the negative binomial, allowing us to derive an estimator which is conditional on the total number of patents applied for by the firm in all the observed years.

This second approach was described in our earlier paper [Hausman, Hall, and Griliches 1984]; by conditioning on the total number of patents applied for by the firm, it essentially allows for a different intercept for each firm. Due to the multiplicative nature of the error in this model, this translates into a different variance for each firm, so that the conditional model estimates an overall variance parameter, but not the individual intercepts or variances. The log likelihood for this model is

$$\log L = \sum_i \sum_t \log \Gamma(\lambda_{it}+y_{it}) - \log \Gamma(\lambda_{it}) - \log \Gamma(y_{it}+1) + \log \Gamma(\sum_t \lambda_{it}) + \log \Gamma(\sum_t y_{it}+1) - \log \Gamma(\sum_t \lambda_{it} + \sum_t y_{it}) \tag{6}$$

Table 6 gives the results of estimates obtained in both ways for both of our samples of firms. The first two columns are estimates of the conditional negative binomial model, while the last two are estimated using weighted nonlinear least squares on equation (1), where the firm effect s_i includes all the R and D variables in all years, but with coefficients δ constrained to be the same across the years. These two methods of estimation are both compromises of a different sort: the negative binomial version allows for an arbitrary firm effect while making a specific

TABLE 6
ESTIMATES WITH FIRM EFFECTS

Number of Firms	Conditional Negative Binomial 642	346	GMT with Correlated Effects 642	346
Log R_0	.29 (.04)	.32 (.07)	.23 (.07)	.30 (10)
Log R_{-1}	−.01 (.05)	−.08 (.09)	−.02 (.07)	−.10 (.08)
Log R_{-2}	.08 (.06)	.06 (.09)	.04 (.06)	.06 (.06)
Log R_{-3}	.02 (.04)	−.01 (.01)	.03 (.06)	−.0005 (.06)
Log R_{-4}		.04 (.07)		.06 (.07)
Log R_{-5}		.01 (.05)		.04 (.07)
Sum log R	.38	.33	.29 (.08)	.36 (.12)
Log likelihood	−131,539.	−96,362.		

All equations contain time dummies.

[6] The extension of Chamberlain's approach to the nonlinear case is only approximate. In order for this approach to be valid here, we need that the expectation of the fixed effect conditional on log R and D be linear. This will be true, for example, if the independent variables are jointly normal.

distributional assumption while the GMT version controls for a firm effect correlated with R and D in a particular way (linear in the exponential function) but does not impose a distribution on the error term. It is therefore reassuring that the differences between them are not huge.

The basic result is that none of the coefficients are significant except those on current R and D, although the total effect of the lagged R and D does seem to add about .05 to the coefficient on the sum. It makes very little difference whether we look at the 642 firm sample or the sample of 346 firms which has a longer R and D history. From the fact that the coefficient on contemporaneous R and D hardly changes from the unconditional estimates, while that on the last R and D goes to zero we conclude that most of what we have removed by conditioning is the R and D prior to our longest lag. This confirms the result of Table 5 where we saw a considerable smearing of the lag coefficients when we used a longer lag in the unconditional estimates. However, the coefficients are fairly unstable and the standard errors are large, so the most we can say is that there appears to be a fairly strong contemporaneous effect, even when firm effects are controlled for. Evidence for a contribution of lagged R and D to current patenting activity is of the order of about 0.05 in the conditional estimates and possibly larger in the unconditional.

Using an idea in Pakes and Griliches [1984b], we can try to estimate more lags in this equation by assuming that R and D follows a low order AR process, in this case AR1. Since this implies a correlation only between the last included R and D and the presample R and D, the estimates of all coefficients except the last will be unbiased by the omission of earlier R and D. Accordingly, we leave the last coefficient free in each year of the equation, which allows us to estimate six lags in the 1972–1979 sample and eight lags in the 1970–1979 sample. The precision of the estimates declines with the length of the lag since we have fewer and fewer observations for the longer lags (lag six in the 1972–1979 sample is estimated only from the 1979 equation, for example). However, this constraint allows us to use all but one of the years of data on patents for each sample of firms, so that we have seven years in the 642 firms sample and nine in the 346 firms sample.

We show these results in Table 7; they are essentially the same as the conditional estimates in Table 6. We also estimated this version of the model including firm effects correlated with R and D; these turned out to be insignificant ($X^2(7)=3.9$ for the first column and $X^2(9)=11.2$ for the second), although the model in this form is highly collinear so that it is difficult to draw firm conclusions.

The basic message of the results in this section is that permanent differences across firms in the propensity to patent do not appear to bias our estimates of the distributed lag relationship between patenting and R and D, except insofar as they are related to the presample history of R and D. The results of the previous section suggested that this bias, if it existed, would appear of equal magnitude in the first and last lag coefficients of R and D. This does not seem to be the case; what bias there is seems only to affect the last lag, and is eliminated by modelling the differences in the propensity to patent across firms: the size variable and the

TABLE 7

GMT ESTIMATES ASSUMING ARL FOR R & D

Number of Firms years	642 1972–1979	346 1970–1979
Log R_0	.33 (.09)	.26 (.09)
Log R_{-1}	.03 (.06)	.04 (.04)
Log R_{-2}	.05 (.06)	.01 (.04)
Log R_{-3}	−.03 (.06)	−.03 (.05)
Log R_{-4}	.11 (.08)	.09 (.06)
Log R_{-5}	−.08 (.10)	−.04 (.06)
Log R_{-6}	.001 (.19)	.05 (.08)
Log R_{-7}		.17 (.12)
Log R_{-8}		−.27 (.14)
Sum log R	.41 (.24)	.29 (.19)
Log book plant in 1972	.18 (.04)	.20 (.05)
Dummy (sic. sector)	.20 (.10)	.27 (.12)

scientific sector dummy are still significant in the final version of the model. Nor does it imply that all the differences are uncorrelated with R and D, but only that the correlation which is observed can be successfully explained by controlling for the part of the R and D history which we do not observe.

6. CONCLUSION

What do we conclude from this lengthy exploration of a basically simple model? First, there does seem to be a rather strong contemporaneous relationship between R and D expenditures and patenting, which does not disappear when we control for the size of the firm, its permanent patenting policy, or even the effects of its R and D history. The remaining elasticity appears to be about .3 with a fairly large standard error. Second, the contribution of the observed R and D history to the current year's patent applications is quite small, on the order of .05. Third, the contribution of the unobserved or presample R and D appears to be large, about .25, and is a possible explanation of the existence of the observed differences across the firms in the propensity to patent.

One of the most interesting results in this paper has nothing to say about patenting, although it provides one reason why we have difficulty measuring the relationship within a firm over time: the characterization of the pattern of R and D investment within a firm as essentially a random walk with an error variance which is small (about 1.5 percent) relative to the total variance of R and D expenditures between firms. In other words, R and D budgets over this short horizon (eight years) are roughly constant or growing slightly (in constant dollars) and therefore it is difficult to estimate complicated lag structures in the presence of such high multicollinearity. The R and D series seem quite smooth within firms primarily because we are comparing firms of vastly different size; the total standard deviation of the log of R and D expenditures for our sample is about 3.8, whereas it is only 0.3 within firms. But this still corresponds to changes on the order of twenty percent per year in R and D investment. We intend to explore further the within firm pattern and determinants of R and D investment in future work.

Finally, it is difficult to give a clear cut answer to the question this paper was originally designed to answer: is there a significant longrun effect of successful R and D investment in the sense of contributions to advances in knowledge for a number of years in the future for which current patents could serve as an indicator? The evidence presented here indicates that the longrun level of R and D can be quite important, but the result is predicated on inference about the unobserved past of the R and D process. There is very little direct evidence of anything but simultaneity in the year-to-year movements of patents and R and D. This finding suggests another way of looking at the process: in large industrial firms the fraction of R and D expenditures devoted to development rather than basic or applied research tends to be well over fifty per cent (NSF 1982). It seems reasonable to suppose that successful research leads both to a patent application and to a commitment of funds for development. A detailed investigation of this timing is beyond the scope of annual data, but the strong evidence of simultaneity in patents and R and D in our data conforms very well to this picture.

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 and D to patents lag structure better, our overall findings are quite interesting, showing a persistent significant contemporaneous relationship of R and D and patenting and rather wide and semi-permanent differences across firms in their patenting and R and D policies. The later finding provides the challenge for further research in a different style: trying to understand how and why firms differ in their responses to the technological environment they find themselves in.

REFERENCES

ABEL, ANDREW B., " R and D and the Market Value of the Firm: A Note", in, Z. Griliches (ed.), *R and D, Patents, and Productivity*, (Chicago: University of Chicago Press, 1984), 261–270.

BHARGAVA, A. AND D. SARGAN., "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," *Econometrica* 51 (183), 1635–1660.

BOUND, J., C. CUMMINS, Z. GRILICHES, B. H. HALL, AND A. JAFFE., "Who Does R and D and Who Patents?", National Bureau of Economic Research Working Paper #908, in Z. Griliches, ed., *R and D, Patents, and Productivity*, (Chicago: University of Chicago Press, 1984), 21–54.

CHAMBERLAIN, GARY., "Multivariate Regression Models for Panel Data," *Journal* of *Econometrics* 18 1, (1982) 5–46.

————., "Omitted Variable Bias in Panel Data: Estimating the Returns to Schooling," *Annales de l'Insee*, #30–31: (1978), 49–82.

CUMMINS, C., B. H. HALL, AND E. LADERMAN., "The R and D Master File: Documentation," (1983), unpublished.

EICKER, F., "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, 1967).

GOURIEROUX, C., A. MONFORT, AND A. TROGNON, "Pseudo Maximum Likelihood Methods: Application to Poisson Models," *Econometrica* 52 (1984), 701–720.

GRILICHES, Z., "Market Value, R and D, and Patents," *Economics Letters* 7 (1981), 183–187.

HALL, BROWYN H., "Market Value, R and D, and Patents Revisited, or Where Do We Go from Here?", (1985), unpublished.

HAUSMAN, J. A., B. H. HALL, AND Z. GRILICHES, "Econometric Models for Count Data with an Application to the Patents R and D Relationship," *Econometrica* 52 (1984), 909–938.

LEAMER, EDWARD E., *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, (New York: John Wiley and Sons, Inc., 1978).

MACURDY, T., "Time Series Models Applied to Panel Data," mimeo, Stanford University, (1981).

MANSFIELD, E., *Industrial Research and Technological Innovation*, (New York: W. W. Norton 1968).

———, *Research and Innovation in the Modern Corporation*, (New York: W. W. Norton, 1972).

PAKES, ARIEL., "Patents, R and D, and the Stock Market Rate of Return," *Journal of Political Economy* 93 (1985), 390–409.

———., "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," National Bureau of Economic Research Working Paper #1606, (1984), *Econometrica*, forthcoming.

PAKES, A. AND Z. GRILICHES., "Patents and R and D at the Firm Level: A First Look," National Bureau of Economic Research Working Paper #561, iS Z. Griliches, ed., *R and D, Patents, and Productivity*, (Chicago: University of Chicago Press, 1984a), 55–72.

———, "Estimating Distributed Lags in Short Panels with an Application to the Specification of Depreciation Patterns and Capital Stock Constructs," *Review of Economic Studies* LI (2) (1984b), 243–262.

SCHANKERMAN, M. AND A. PAKES., "Estimates of the Value of Patent Rights in European Countries During the Post 1950 Period," National Bureau of Economic Research Working Paper #1650, (1984).

STONEMAN, PAUL., "Patents and R and D: Searching for a Lag Structure — Comment," *Proceedings of the Conference on Quantitative Studies of Research and Development in Industry*, (Paris: ENSAE, 1983).

TAYLOR, C. T. AND A. SILBERSTON., *The Economic Impact of the Patent System*, (Cambridge: The University Press, 1973).

WHITE, HALBERT., "Nonlinear Regression on Cross-Section Data," *Econometrica* 48 (1980), 721–746.

———., "Instrumental Variables Regression with Independent Observations," *Econometrica* 50 (1982), 483–500.