NBER WORKING PAPER SERIES

PATENTS AND R & D AT THE FIRM LEVEL: A FIRST LOOK

Ariel Pakes

Zvi Griliches

Working Paper No. 561

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

October 1980

This paper was presented at the NBER's 1980 Summer Institute in Productivity. This work has been supported by the NBER Productivity and Technical Change Studies Program and NSF grants PRA 79-13740 and SOC 79-04279. We are indebted to M. I. Nadiri for access to unpublished data, to John Bound and Browyn Hall for research assistance, and to Jacques Mairesse for several valuable comments. The research reported here is part of the NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #561
October, 1980

PATENTS AND R&D AT THE FIRM LEVEL: A FIRST LOOK

Abstract

This is a first report from a larger study of inventive activity of U.S. firms and some of its consequences. It reports on the relationship between patents applied for and R&D expenditures based on data for 121 large corporations covering the 1968-1975 period. The main conclusion is that there is a statistically significant relationship between a firm's R&D expenditures and the number of patents it applies for and receives. This relationship is very strong in the cross-sectional dimension (squared partial correlations of .8 or higher). It is weaker in the within-firm time-series dimension (partial r^2's of .2 to .3). Attempts to fit an unconstrained distributed lag relationship yields only significant coefficients for the first and last terms in the lag structure, indicating both a quick response of patenting to changes in R&D and a small but persistent effect of past R&D, the truncation of this long lag being reflected in a significant coefficient for R&D lagged five years. In spite of these difficulties, patent counts do measure something systematic and hence are worthy of further study.

Ariel Pakes
Zvi Griliches

National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

(617) 868-3921

Ariel Pakes and Zvi Griliches
NBER and Harvard University
May 1980
First Draft

PATENTS AND R&D AT THE FIRM LEVEL: A FIRST LOOK*

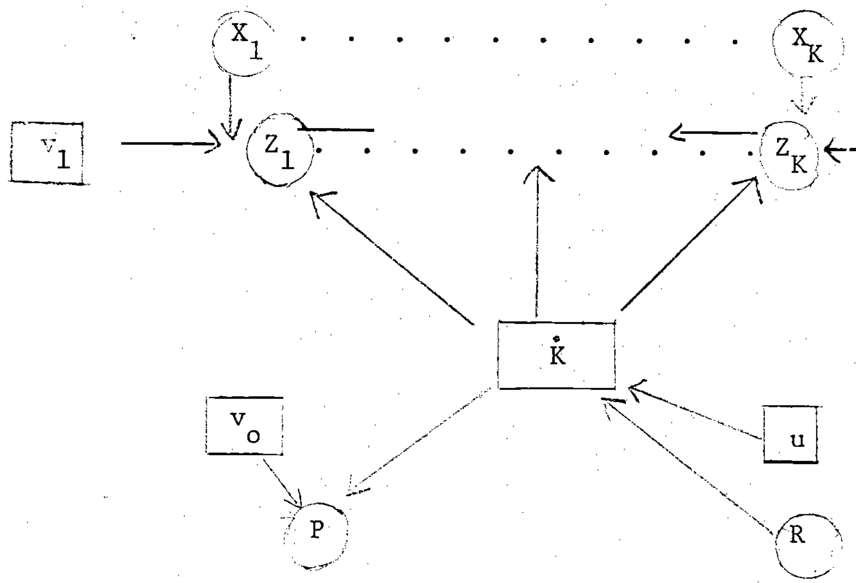
I. Introduction

This paper is the first report from a more extensive study of knowledge producing activities in American industry initiated by the National Bureau of Economic Research. Perhaps the most serious task facing empirical work in the area of "technological change" and "invention and innovation" is the construction and interpretation of measures (indices) of advances in knowledge.[1] If one defines K as the level of economically valuable technological knowledge, and $\dot{K} = dK/dt$ as the net accretion to it per unit of time, then the first task of our research program is to evaluate the usefulness of several indicators of $\dot{K}$, focusing particularly on patents and the value of the firm, variables which have yet to receive the attention that we think might be warranted in this context.[2]

[1] For a thoughtful discussion of this point see Kuznets (1962).

[2] Most of the previous work on patents is either quite ancient and/or inconclusive. Professional opinion has not really progressed much past the disagreement about the utility of patent statistics reflected in the discussions between Kuznets, Sanders and Schmookler [Nelson (1962)]. The most recent review of the literature and independent contribution is to be found in Taylor and Silberstone (1978). The papers that come closest to the topics treated here are Scherer (1965) and Comanor and Scherer (1969).

A simplified path analysis diagram of the overall model. Squares denote unobservable magnitudes and circles denote observable ones.



The Z's are indicators of other expected or realized benefits from invention

The K.P.F.

Quantitative Indicator of Number of Inventions

Research Expenditures

FIGURE 1

The basic structure of our project is illustrated succinctly by the path analysis diagram in Figure 1. In that diagram $\dot{K}$ is a central unobservable which together with the observable X's and the disturbances, the v's, determines the magnitude of several interrelated indicators of invention and innovation, the Z's. The latter include the stock market value of the firm, the productivity of traditional factors of production, and investment expenditures on traditional capital goods. It is an investigation of the lower half of this diagram which we report on in this paper. We see there that $\dot{K}$ is produced by a knowledge production function, K.P.F., which translates past research expenditures, R, and a disturbances term u, into inventions. The disturbance term reflects the combined effect of other non-formal R&D inputs and the inherent randomness in the production of inventions. Patents, P, are an imperfect indicator of the number of new inventions with v_o representing the noise in the relationship between P and $\dot{K}$. It is clear from the figure that the patent equation, the equation connecting patents to past research expenditures, combines the properties of both the K.P.F. and of the indicator function relating to P and $\dot{K}$. Without additional indicators of $\dot{K}$ one cannot separate the two types of effects. For example, both u and v_o enter the relationship between R and P but only u affects the Z's. In the context of a larger model, one could separate out the effects of u from v_o by calculating the effect of the residual in the patent equation on the Z's, but this cannot be done from the patent equation alone.

We have made several simplifications in drawing and discussing this diagram. For example, the relationship between K and $\dot{K}$ should be defined explicitly to allow for the possibility of decay in the private value of knowledge. $\dot{K}$ may be determined by the absolute level of K as well as by past investments in research resources. If, as is likely, the u's are correlated over time then one would expect any realization of u to feedback into the demand for research resources. Moreover, conditions (economic, technological and legal) should be specified under which the benefits from applying for a patent outweigh the costs of the patenting process, adding thereby more structure to the relationship between P and $\dot{K}$.[3] Figure 1 does, however, provide an overview of our project and is sufficiently precise for the discussion of the two issue on which this paper will concentrate:

1) the "quality" of patent counts as an indicator of knowledge increments and;

2) The time-shape of the lag between research expenditures and patentable results.

[3] Such a theory, we think, would be based on the underlying notion of a research project whose success depends stochastically both on the amount of resources devoted to it and on the amount of time that such resources have been deployed. With each technical success there is associated an expectation of the ultimate economic value of a patent to the inventor or his employer. If this expectation exceeds a certain minimum, the cost of patenting, a patent will be applied for. That is, the number of patents applied for is a count of the number of successful projects (inventions) with the economic value of a patent exceeding a minimal threshold level. If the distribution of the expected value of patenting successful projects remains stable, and if the level of current and past R&D expenditures shifts the probability that projects will be technically successful, an increase in the number of patents can be taken as an indicator of an upward shift in the distribution of $\dot{K}$. Whether the relationship is proportional will depend on the shape of the assumed distributions and the nature of the underlying shifts in them. What we are dealing with here is at best a very crude reduced form type equation whose theoretical underpinnings remain still to be worked out. But one has to start someplace.

The recent computerization of the U. S. Patent Office's data base has made it possible for the first time to follow the patenting behavior of a large cross-section of firms over a significant time interval. This makes patent counts an easily accessible, perhaps the most easily accessible, indicator of the number of inventions made by a firm. Moreover, patents are a quantitative and rather direct indicator of invention; an indicator which is not contaminated by many of the X's which also affect the Z's . There are, however, several problems with the patent measure, the major ones being the fact that not all new innovations are patented and that patents differ in their economic impact. These considerations have led to doubts about the "quality" of patent counts as an indicator of knowledge increments (see the literature cited in footnote 2). We attempt to respond to such concerns by first presenting a more precise description of the patent equation in section II of this paper and then reporting in section III on one particular measure of the "quality" of patent statistics.

There is another advantage of patent counts over other indicators of knowledge production. Patents are applied for at an intermediate stage in the process of transforming research input into benefits from knowledge output. They can be used, therefore, to separate the lags that occur in that process into two parts; one which produces patents from current and past research investments, and another which takes patents, with the possible addition of more research expenditures, into benefits. Such a breakdown should allow us to estimate more precisely the overall lag structure, a structure which has confounded and confused previous empirical work in this area.[4] Section IV presents our first-round estimates of the distributed lag between research expenditures and patentable results.

The data used in this study are at the firm level and are based on a merger of the information provided in the Compustat file (based on the 10-K firm reports to the SEC) and patent data tabulated by the Office of Technology Assessments and Forecasts of the U.S. Patent Office. These data and the particular sample chosen are described in greater detail in Appendix A. Most of the work reported on here is based on the patenting behavior of 121 firms during the 1968-75 period.

[4] See, for example, how two different assumptions about the lag structure lead to very different calculations of the private rate of return to research expenditures from the NSF-Griliches data; Z. Griliches (1980) vs. A. Pakes and M.A. Schankerman (1978).

Section II: The Model

We report, in Appendix B, a preliminary investigation into the functional form of the relationship between patents and past R&D expenditures. That analysis supports a rather simple patent equation: the logarithm of patents (p) as a function of a time trend (t), current and five lagged values of the logarithm of research expenditures (r) and a set of firm specific dummy variables. In this section we provide an interpretation of this patent equation in terms of a simple model relating past r to the logarithm of current knowledge increments ($\dot{k}$), and $\dot{k}$ to p.

Consider first the transformation function from r to $\dot{k}$, or the K.P.F. Assuming it to be of the Cobb-Douglas form but allowing for firm constants and a time trend, we have:

$$\dot{k}_{i,t} = a_i + bt + \sum_{\tau=0}^{5} \theta_\tau r_{i,t-\tau} + u_{i,t} \qquad (1)$$

where $u_{i,t}$ is an independent and identically distributed disturbance which is not correlated with r and represents randomness in the K.P.F.

The a_i represent firm-specific differences in the productivity of research effort caused by either variance in technological opportunities or differences in managerial ability. Such differences will, in general, be transmitted to differences in research expenditures, firms with more productive research departments investing more in research. Thus, the a_i have two roles in the subsequent analysis. First, they cause differences in $\dot{k}$ and this should be considered in an analysis of the determinants of the variance in p. Second, their correlation with the $r_{t-\tau}$ must be accounted for in any attempt to estimate the θ_τ or else the coefficient estimates will be a combination of the effect of the $r_{t-\tau}$ on $\dot{k}$ (the θ_τ) and the effect of a_i on r. To be more explicit about the latter point we simply project the a_i on all in-sample research expenditures. Since the a_i are constant over time they can only be correlated with the means of the research variables. We can write, therefore,

$$a_i = \sum_{\tau=0}^{5} \phi_\tau r_{i.,-\tau} + u_i. \qquad (2)$$

where

$$r_{i\cdot 0} = T^{-1} \sum_{t=1}^{T} r_{it} \quad , \quad r_{i\cdot -1} = T^{-1} \sum_{t=0}^{T-1} r_{it-1} \quad , \quad \text{etc.,}$$

and u_i is by construction uncorrelated with all in-sample research variables.[5]

[5] In econometric terminology the model we are working with is a variant of the partial transmission model of Mundlak and Hoch (1965). The unobservable portion of the K.P.F. which is transmitted to the research demand equation is assumed to remain constant over time. This assumption, plus the nature of the panel, will allow us to use single equation estimation techniques to estimate parameters of the patent production function. A more precise discussion of the econometric techniques underlying the estimation procedures to be used in this paper is to be found in Mundlak (1978) and Pakes (1978, chapter 3).

Patents are our indicator of knowledge increments. If one allows for a time trend in the relationship between p and $\dot{k}$ that relationship is written as:

$$p_{i,t} = dt + \beta\dot{k}_{i,t} + v^{*}_{i,t} \qquad (3)$$

where v^{*} is uncorrelated with $\dot{k}$ and t by construction.

(3) should be interpreted as a reduced form from the appropriate patenting model. In that reduced form β is the elasticity of patents with respect to knowledge increments and d is a measure of the trend in factors determining the propensity to patent. $v^{*}_{i,t}$, on the other hand, is that part of the (detrended) variance in patents which cannot be accounted for by (detrended) movements in knowledge increments; that is, variance in $v^{*}_{i,t}$ is "noise" in the patent measure. To facilitate interpretation we will make two assumptions on $v^{*}_{i,t}$. First, we let $v^{*}_{i,t}$ be composed of a firm-specific component, v_i, which reflects differences among firms in their <u>average</u> propensity to patent, and a second, independent and identically distributed disturbance, $v_{i,t}$, reflecting the <u>variations</u> (around a trend) in the propensity to patent of a given firm over time. Thus, $v^{*}_{i,t} = v_i + v_{it}$. Second, since $v^{*}_{i,t}$ is uncorrelated with $\dot{k}_{i,t}$ (by choice of β) we shall assume also that its determinants v_i and $v_{i,t}$, are each uncorrelated with the determinants of $\dot{k}$ (the r's and u's) given

by (1) and (2).[6]

Substituting (1) and (2) into (3) we can now provide an interpretation to the equation preferred in our analysis of functional form, that is to the equation:

$$p_{i,t} = \alpha + \gamma t + \sum_{\tau=0}^{5} w_\tau r_{i,t-\tau} + \sum_{\tau=0}^{5} \phi_\tau r_{i.,-\tau} + \eta_i + \varepsilon_{i,t} \tag{4}$$

where

$$w_\tau = \beta\theta_\tau \quad , \quad \gamma = \beta b+d$$

$$\eta_i = v_i + \beta u_i \quad , \text{ and } \quad \varepsilon_{i,t} = \beta u_{i,t} + v_{it}$$

for

$$i = 1,\ldots,N \quad \text{and} \quad t = 1,\ldots,T.$$

[6] The first assumption allows us to provide standard errors for our estimates of the regression coefficients. The second is a rather strong assumption. We are assuming that randomness in the K.P.F., above or below average success in converting research expenditures into knowledge increments, does not influence the patenting decision, that the two sources of randomness are distinct and independent. We need this assumption to make the interpretations that follow.

The first point to note from (4) is that though one cannot estimate the elasticities of knowledge increments with respect to research resources, the θ_τ , one can investigate the form of the distributed lag connecting $\dot{k}$ and r , since $w_\tau/\Sigma w = \theta_\tau/\Sigma\theta$. The sum of the estimate lag coefficients, $w^* = \sum_{\tau=0}^{5} w_\tau$, estimates the product of the degree of economies of scale in the K.P.F., $\sum_{\tau=0}^{5} \theta_\tau$, and the elasticity of patents with respect to knowledge increments (β). These two parameters can be identified separately only in a larger model which includes additional indicators of the benefits from knowledge producing activities (see the Introduction).

Recall that the various variance components which combine to form the disturbance term in (5) are mutually uncorrelated. It follows that $\mathrm{Var}\ (\eta_i + \varepsilon_{i,t}) = \sigma^2$, the variance of the total disturbance in the patent equation is greater than $\mathrm{Var}(v_i+v_{i,t})$, the variance of the noise in patents as an indicator of $\dot{k}$. It also implies that, temporarily ignoring the time trend in the patent indicator equation (assuming $d = 0$), the ratio of σ^2 to the total variance in the logarithm of patents $(1-\bar{R}^2)$ provides an upper bound for the noise to total variance ratio in the patent measures. The upper bound will be called $\lambda^{u.T.}$ and its complement, the relevant $\bar{R}^2$ measure, is a measure of the quality of patents as an indicator of knowledge increments. If, instead of assuming $d = 0$, we assume $b = 0$, that is, the entire trend effect is due to differences in the average propensity to patent over time, then one can derive an analogous measure of $\lambda^{u.T.}$ for detrended patents by filtering out time from both the patent and R&D variables. In practice, the two measures of $\lambda^{u.T.}$ were always almost identical. In the next section, we also present the comparable information on the noise to total variance ratio in the between firm variance in patents (i.e., in the variance of $P_{i\cdot} - P_{\cdot\cdot}$), labelled λ^{uB}, and in the within firm variance in patents (the variance in $P_{it} - P_{i\cdot}$) $\lambda^{u.W.}$. The latter two statistics provide some indication of the usefulness of patent counts as an indicator of knowledge increments for studies of invention and innovation that focus either on cross-sectional differences in the production of knowledge between firms, or on the within firm fluctuations over time.

Section III: Measures of the Quality of the Patent Variable

Table 1 presents estimates of $1-\lambda^{u.T.}$, $1-\lambda^{u.W.}$, $1-\lambda^{u.B.}$, the lower bounds to the systematic to total variance ratios, σ_{ε}^{2}, σ_{η}^{2}, and some relevant sample moments for each of the seven industries in our data (rows 0 through 6), all firms in our sample (row 7) and firms in the industries defined by rows 1 through 6 (row 8). The latter sample concentrates on firms in research-intensive industries.

Starting with the measures of $1-\lambda^{u.T.}$ in the separate industries, it is clear, even from our simplistic model, that much of the patent variance is systematic, providing a good indicator of the underlying variance in $\dot{k}$. For the seven industries in our sample, about 85% of the variance in p is associated with variance in r and, in some industries, notably scientific instruments and office computing and accounting machinery, the lower bound of the systematic to total variance in patents is closer to .95.

These estimates hide, however, some relevant information. Moving to column (2) it is clear that we are far less certain of whether within any given firm changes over time in p reflect systematic changes in knowledge production by that firm. In the within firm calculations it mattered whether or not we first filtered out time trends from p and r. Therefore, the bracketed numbers beside column (2) refer to systematic to total variance ratios in detrended patents. Averaging over the seven industries we find that the lower bound $(1-\lambda^{u.W.})$ is only around 20 to 25%, though it does reach 50% in office, computing and accounting machinery. Without the larger model alluded to in the introduction one cannot really tell whether the smaller systematic to total variance ratios in the "within"

Table 1: Lower Bounds to the Systematic Variance Ratio in p and Some Sample Moments for the Seven Industries

	Industry Description[d]	Lower Bound to the Systematic Variance ratio[a] $1-\lambda^{u.T.}$ total	$1-\lambda^{u.W.}$ within	$1-\lambda^{u.B.}$ between	σ_{ε}^{2}[b]	σ_{η}^{2}[c]	Variance in p	Ratio of Within to Total Variance in p	Variance in r	Ratio of Within to Total Variance in r	Firms N	Observations NT
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
0	Other manufacturing	.74	.19 (.16)	.77	0.17	0.50	2.71	0.08	1.83	0.04	41	328
1	Industry 28 except 283 (Chemicals & Allied Products except Drugs and Medicines)	.82	.13 (.11)	.86	0.10	0.26	2.02	0.06	1.16	0.02	19	152
2	Industry 283 (Drugs and Medicines)	.80	.33 (.22)	.85	0.07	0.14	0.85	0.10	0.56	0.03	19	152
3	Industry 35 except 357 (Machinery except Office, Computing and Accounting)	.82	.11 (.06)	.87	0.16	0.32	2.97	0.07	2.17	0.04	13	104
4	Industry 357 (Office, Computing and Accounting)	.84	.52 (.50)	.96	0.16	0.09	3.78	0.11	2.06	0.11	10	80
5	Industry 366/367 (Electronic Components & Communications)	.51	.46 (.42)	.89	0.07	0.07	0.83	0.13	1.57	0.03	8	64
6	Industry 38, (Professional & Scientific Instruments)	.95	.28 (.06)	.97	0.08	0.05	2.55	0.04	2.21	0.04	11	88
7	Total Sample	.66	.33 (.23)	.69	0.14	.66	2.41	0.07	1.72	0.04	121	968
8	Firms in Research Intensive Industries	.73	.33 (.22)	.76	.12	.48	2.20	0.07	1.61	0.04	80	640

[a] The bracketed numbers refer to lower bounds to the systematic variance ratio in patents after a time trend has been filtered out of both p and r. p - log patents. r - log R&D.

[b] Calculated as the variance of estimate from an O.L.S. regression of $p_{i,t}$ on current and five consecutive lagged years of $r_{i,t}$ and firm specific constants.

[c] Calculated as the variance of estimate from an O.L.S. regression of equation (4) in the text minus the estimate of σ_{ε}^{2}

[d] Two and three digit industry identification numbers refer to SIC codes.

data reflect true randomness in the knowledge production function, small differences in research expenditures over time within a given firm having very sporadic effects on the production of inventions in particular years, or whether they arise because firms decide to patent different proportions of their inventions in different years.

Two more points should be noted about the results for the separate industries. Column 6 shows that over 90% of the total variance in p is between firm variance. As a result $1 - \lambda^{u.B.}$ is very close to $1 - \lambda^{u.T.}$. Second, though σ_{ε}^{2} does not vary too much between the sample industries, σ_{η}^{2} varies a lot, being much larger in the less homogeneous industries (rows 0, 1, and 3). This is likely to reflect greater differences in the average propensities to patent in those industries.

Looking at the samples which aggregate the various industries (rows 7 and 8) we find that the $1 - \lambda^{u.W.}$ actually decrease after pooling different industry samples. This implies that, at least in our sample, the elasticity of patents with respect to knowledge increments (β) and the response of $\dot{k}$ to current and past r (θ_{τ}) do not vary much between the industries aggregated; a result which will be confirmed in the next section.

Section IV: Coefficient Estimates

Table 2 presents the estimates of the w_τ and the coefficient of the trend term based on data from all of the 121 firms and estimates based on two sub-samples: firms in research-intensive industries and "other manufacturing" firms. Row 10 of that table presents the estimates value of the F statistic for the null hypothesis that these coefficients do not differ between the industries aggregated. The tests statistics indicate that, after we allow for a separate trend and intercept for the drug industry (row 9), our sample cannot really pick up any additional interindustry differences in coefficients.

Turning to the coefficient of the trend term, that coefficient was negative, and significantly so, for all industries except for the drug industry. There are two alternative explanations of this result and they cannot be separated out without the larger model alluded to in the introduction. First, the negative trend is consistent with impressionistic evidence on the declining propensity to patent in U.S. manufacturing. The drug industry is indeed an exception since, during the period concerned, there occurred both a relaxation in the patent office's acceptance procedures regarding patents on natural substances and significant changes in regulatory

[7] The possible exception here is the drug industry. When that industry was dropped from the first two samples the observed values of the F test dropped significantly to 1.37 and 1.67 respectively. Still the estimated coefficients for the drug industry were not very different from those of the other industries in the sample, except for the trend coefficient.

Table 2: Distributed Lag Estimates*

		(1)	(2)	(3)
		All Firms	Firms in Research Intensive Industries	Other Manufacturing Firms
1.	r_o	.56 (.07)	.52 (.10)	.62 (.14)
2.	r_{-1}	-.10 (.09)	-.01 (.12)	-.22 (.16)
3.	r_{-2}	.05 (.09)	.08 (.12)	-.02 (.16)
4.	r_{-3}	-.04 (.09)	-.21 (.13)	.13 (.15)
5.	r_{-4}	-.05 (.10)	-.01 (.15)	-.08 (.16)
6.	r_{-5}	.19 (.08)	.25 (.11)	.13 (.14)
7.	Sum (w^*)	.61 (.08)	.62 (.09)	.61 (.04)
8.	t	-.04 (.007)	-.05 (.008)	-.03 (.012)
9.	t_{Drugs}	.07 (.01)	.07 (.01)	
10.	F aggregation (critical values, 1%, 5%)	1.54 (1.39, 1.58)	2.08 (1.45, 1.69)	
11.	Degrees of Freedom	837	550	279

* Standard errors are in brackets below coefficient estimates.

conditions facing that industry.[8] The same result, however, could have been caused by a secular decline in the private productivity of research resources, a hypothesis which is consistent with the observed negative growth rate of employment of R&D scientists and engineers during the period considered.[9]

The individual coefficients are not estimated very precisely. The sum of the lags, w^*, is estimated with a fair amount of precision and equals about .60 with a standard error of 0.08. If one ignores the fact that some of the estimated lag coefficients are negative and computes a "mean lag," it equals about 1.6 years for the all firm sample. Unless substantial R&D is done on projects after patents are applied for, this should approximately equal the mean R&D project gestation lag, the lag between project inception and project completion. The scattered empirical evidence on gestation lags indicates that this is indeed the case.[10]

Still, the estimated form of the lag is rather disturbing. There are large, significant, positive coefficients in the first and last years, and very little effect of interim R&D on patent applications. Though the current year's coefficient could indicate the presence of simultaneous

[8] For a description of the effect of these events see P. Temin (1979)

[9] See Griliches (1980) for a similar finding on aggregate data.

[10] Sources of project level data are L. Wagner (1968) and J. Rapoport (1971). This evidence is summarized in terms of mean gestation lags in A. Pakes and M.A. Schankerman (1978). The average of the mean gestation lags presented in the latter paper was 1.34 years.

equations bias, that is not really a necessary implication of the results. The R&D project level data cited above do point to a gestation lag which is highly skewed with large early year coefficients and any minor misspecification in the model could push all this effect into the coefficient of r_0. The coefficient estimate which is perhaps more disturbing is that of the last year since it could be indicating the presence of a "truncation" problem in our distributed lag estimates. That is, the coefficient of the fifth year could be proxying for a series of small effects of the more basic research done six years ago or earlier. There estimates of the form of the lag should be treated with caution, both because of the possible truncation problem and because they are not really consistent with our prior beliefs about the form of this lag structure.

Section V: Conclusions and Extensions

Our first look at the patent equation suggests the following conclusions. First, the data were quite clear on the form of that equation; log-log with (correlated) firm effects and a time trend being preferred over alternatives. Second, our major positive finding is given by the $1-\lambda^{u.B.}$ estimates presented in Table 1. They show that patents are a good indicator of between firm differences in advances of knowledge. Since the between firm component dominates the total variance in patents a similar comment applies also to the total variance. If this result changes at all in the more sophisticated models we are beginning to estimate, it is only likely to improve. Use of a longer series of past R&D expenditures cannot but increase the fit of the patent equation, and adding another indicator of benefits will separate out the effect of randomness in the K.P.F., the u, from the effect of noise in the patent measure, the v^*, allowing us to narrow the bound further.

The rest of our results are not as heartening. While a part of the within firm variance in patents is related to the variance in R&D expenditures, a significant portion, about 75 percent, is not. At this stage we cannot tell whether the fault lies in the patent measure (the variance in v^*), in randomness in the K.P.F. (the variance in u), or in simple errors of measurement in both p and r. Most of the coefficients, except for trend, were not estimates very precisely. This is a result of two factors, First, only the within firm variance in p and r can be used to estimate w_T and this variance is a small part of the total variance in these variables (see Table 1). The second factor leading to imprecise estimates is the small sample size (maximum T = 8, N = 121). We can and will increase our sample significantly in the future by not insisting that firms had to have reported R&D expenditures before 1972 (see Appendix A). Including such firms will force us, however, to use only a few lagged terms of r or assume a specific functional form for the distributed lag between patents and R&D expenditures, even though we have yet to acquire much information on the shape of this distribution. Because our estimates indicate that even with 5 lagged R&D terms we still may have a truncation problem, we have been developing a technique for estimating distributed lags in panel data when the time series on the independent variable is short. We are also investigating the impact of other sources of bias in the estimated coefficients, in particular the effect of measurement errors in the R&D variables. Finally, once an appropriate specification for the patent equation has been determined, we will combine it with the other equations in our model in the hope of providing a fuller understanding of the process of invention and innovation in American industry.

In short, a great deal of work remains to be done, but we have made a start. It is already clear that something systematic and related to knowledge producing activities is being measured by patents and that they are, therefore, very much worthy of further study.

Appendix A: Data Sources, Sample and Variable Definitions, and Sample Characteristics

The data base used in this preliminary round is neither complete nor representative. We have tried to get together from published sources as large a sample of firms as possible, covering the 1963-77 period. The main selection variable is R&D. Until recently (1972 and later) most firms did not report their R&D expenditures publicly. The firms that did report R&D expenditures, reported company financed R&D expenditures, and those numbers are recorded on the Standard and Poor's Computstat tape, which served as a major source of our data.[1] An earlier study by Nadiri and Bitros (1980) had used the Compustat tape and a mail survey to fill in some of the gaps on this tape to construct time series of R&D for 114 firms for the period 1963-72. Starting with a later edition of the tape we found 146 firms with no more than 3 years of R&D data missing during the 1963-75 period. Combining it with the Nadiri sample yielded an unduplicated total of 172 firms. Fifteen firms were eliminated from this total either because they were foreign, had undergone large mergers, or had other unreconcible jumps in their data. This left us with a total of 157 firms which constitute the sample on which we are currently working.[2] For the purposes of this

[1] Only company financed R&D ought to lead to patents since government R&D contracts most often include clauses which put the output of government funded projects in the public domain.

[2] It is possible to construct a much larger sample is one is willing to restrict oneself to post 1972 data. We are currently engaged in developing such a 1972-78 sample, with an expected N of about 900.

paper, the sample was further restricted to firms that had data (did not undergo any major reorganization) throughout the whole period (N = 144) and had an R&D program of more than minimal size (R&D ≥ $0.5 million) in any one year, (N = 121). What we have done, then, basically, is to expand the Nadiri sample slightly, update it to 1977, and add patent data to it.[3]

The patent data were supplied to us by the Office of Technology Assessment and Forecasts of the U.S. Patent Office. They are based on a tape of all patents granted during the years 1969-78. These data are then reclassified by year of application rather than by year of grant. One of our tasks was to be sure that we had all the subsidiaries and names used by a particular corportation. For this purpose we scrutinized the alphabetical index of patenting organization provided by OTAF and checked it against the list of a firm's subsidiaries given in the Dictionary of Affiliations and a list of past mergers given in Mergers and Acquisitions. Where a firm had acquired another one during this period, we added in the patents of the acquired organization (and its R&D expenditures, when known). In a few cases, where the mergers were large and occurred towards the end of the period, we left the two firms unmerged and instead declared the recent (post-merger) years as missing.

Because the patent data are based on patents granted during 1969-78, patents by year applied for cannot really be used before 1968. While only less than 1 percent of all patents granted is granted within the year of application, about 10 percent are granted in the following year. Thus, only about 89 percent of the patents applied for in 1967 would appear among

[3] Some of the missing years have been interpolated by us. Also, the definition of expenditures reported as R&D by different firms may change over time. Where such changes were obvious or stated in the 10-K forms, we tried to adjust for it. Where we could not and the discrepanies were large, we eliminated the firm from our sample.

the patents counted by us. Similarily, one cannot probably use the patent data by year of application after 1975, since it takes about four years after the application before more than 96 percent of the patents applied for in that year that will be eventually granted are actually granted.[4] Thus, at best, we have about 8-9 years of usable patent data. In most of the analyses we used the 8 years 1968-75. Eight years and 121 firms give us an effective sample size of 968.

Table A1 gives means and standard deviations for a few of the major variables in the various samples and industries represented in this study. The industrial classification was chosen to approximate the industrial breakdown used by the NSF in their reports. It is clear from this table that these firms are rather large, that the exclusion of firms with R&D budgets of less than half a million dollars makes them even larger, that the size distribution of the firms is quite skewed (standard deviations are on the order of the means or larger), and that the industrial distribution is quite uneven. The firms represented in the sample are those who reported their R&D expenditures publicly in the 1960's with drug and chemical firms over-represented. In spite of the presence of relatively large firms in this sample, there is still a relatively large number of low patenting firms. Table A2 gives a distribution of firms by their level of patenting.

[4] These estimates are based on an unpublished tabulation of patents granted by date applied for for the period 1965-77 made available to us by OTAF.

Table A1. Characteristics of Sample Firms by Industry: Averages 1963-75 and Standard Deviations

	Entire Sample Ind=0			Firms with Complete Data and Min R&D $\geq$ 500K		
VARIABLE	N	MEAN	STANDARD DEVIATION	N	MEAN	STANDARD DEVIATION
DEFRND	13	5.678	7.291			
GROPLA72	13	211.994	272.923			
PATS	13	15.788	21.781			
		IND=28			IND=28	
DEFRND	21	28.353	32.953	19	31.258	33.362
GROPLA72	21	1053.298	1248.221	19	1162.378	1264.633
PATS	21	92.804	104.502	19	102.520	105.298
		IND=28.3			IND=28.3	
DEFRND	20	26.665	20.009	19	28.013	19.603
GROPLA72	20	264.486	206.233	19	277.698	203.002
PATS	20	54.531	40.089	19	57.316	39.151
		IND=35			IND=35	
DEFRND	14	17.143	25.603	13	18.407	26.191
GROPLA72	14	327.631	429.287	13	352.300	436.366
PATS	14	47.464	64.881	13	50.990	66.119
		IND=35.7			IND=35.7	
DEFRND	13	25.422	30.165	10	32.169	31.521
GROPLA72	13	544.321	767.480	10	665.861	843.293
PATS	13	62.490	98.328	10	79.912	106.895
		IND=36			IND=36	
DEFRND	15	15.457	34.068	8	28.427	43.694
GROPLA72	15	393.137	1248.032	8	731.354	1683.749
PATS	15	37.975	68.848	8	70.031	83.442
		IND=38			IND=38	
DEFRND	15	25.507	46.550	11	34.452	51.996
GROPLA72	15	352.326	815.241	11	477.772	930.342
PATS	15	63.592	99.897	11	85.920	109.149
		IND=99			IND=99	
DEFRND	46	20.489	29.581	41	22.884	30.500
GROPLA72	46	3074.459	10476.016	41	3445.557	11052.898
PATS	46	56.068	90.813	41	61.808	94.687
		Combined			Combined	
DEFRND	144	22.612	30.994	121	26.709	32.228
GROPLA72	144	1331.104	6044.733	121	1578.299	6569.299
PATS	144	59.854	84.891	121	70.565	88.661

DEFRND -- Deflated R&D expenditures

GROPLA72 -- Book value of gross plant in 1972

PATS -- Number of patents, by year applied for.

Footnotes to Table A1

Ind=0 -- Firms with incomplete data for the period as a whole

Ind=28 -- Chemicals and Allied Products except Drugs and Medicines

Ind=28.3 -- Drugs and Medicines

Ind=35 -- Machinery except Office Computing and Accounting Machinery

Ind=35.7 -- Office, Computing and Accounting Machinery

Ind=36 -- Electronic Components and Communications

Ind=38 -- Professional and Scientific Instruments

Ind=99 -- Other Manufacturing

The R&D expenditures have been deflated by an R&D "deflator" index constructed along the lines suggested by Jaffe in NSF (1972). It is simply a weighted average of the index of hourly labor compensation and the implicit deflator in the non-financial corporations sector, with .49 and .51 as relative weights.

The main problem with our sample is its peculiar nature. It is based on those companies that reported R&D expenditures in the mid-60's. Since it is selected on the "independent" variable in this study, one need not anticipate much of a selectivity bias in equations where patents or the market value of the firm are the dependent variables. Moreover, since most of our analysis will be "within" firms, any fixed selectivity adjustment would be incorporated in the constant term and would not affect our inferences. We do intend, however, to explore the selectivity issue later on, once we have completed the construction of a larger sample based on 1972-79 data. Since data are available for most firms as of 1976, one can find out the characteristics of the firms that did not report R&D earlier and adjust our estimates accordingly.

Table A2 Distribution of Sample Firms by Number of Patents: (a) Annual, by year applied for, 1968-75; (b) 10 year average, by year granted, 1969-78.

Patents/year	by observation 1968-1975	by Firm 1969-1978 Average
0	98	1
>0 ≤1	79	7
>1 ≤2	57	11
>2 ≤5	121	19
>5 ≤10	94	16
>10 ≤20	135	9
>20 ≤50	262	38
>50 ≤100	195	26
>100 ≤200	137	17
>200 ≤400	58	10
>400	19	3
total	8x157=1256	157

Appendix B: The Form of the Patent Equation

Because there was little prior empirical or theoretical research on the R&D to patents relation we began our analysis with an investigation of the functional form of the equation that might connect these two variables in our data.

Functional form questions were examined allowing the parameters of all estimated equations to differ in each of our seven industries and between firms with large and small R&D departments within each industry.[1] That is, fourteen sets of parameters were estimated. The independent variables included in the estimating equations were a set of time dummies, the current and five consecutive lagged values of both the logarithm of R&D expenditures and R&D expenditures per se, and a set of firm-specific dummy variables (constants). To simplify matters we assumed that the appropriate form of the dependent variable was either $\log(P) = p$, or P itself. Hence log-log, semi-log, and linear functions each with firm and time effects, were all special cases of the model with which we started.

A variant of the Box and Cox (1962) procedure was used to choose the form of the dependent variable. It indicated that the logarithm of patents was clearly preferred over the absolute number of patents by the data for each separate grouping, and for the sample as a whole. We then asked the

[1] Small firms were defined, quite arbitrarily as firms whose R&D expenditures over the sample period (from 1963 to 1975) fell below half a million dollars in at least one year. The size breakdown had the effect of separating out the recently born science-based firms from the others in the sample and allowed for the possibility that the characteristics of the K.P.F. differed in the firms with smaller, less established, research departments.

question of whether the parameters of the relationship between p and the independent variables within each industry differed between firms with large and small R&D departments. The test statistic was significant at any reasonable level of significance indicating that the form of the relationship between patents and research expenditures was different for firms with small R&D departments. The 26 firms in the small group were dropped from all the subsequent computations reported in this paper. Next, we wanted to know whether the model could be simplified by assuming either that the coefficients of current and all lagged values of R&D, or that the coefficients of the logarithmic forms of these variables, were all zero. The $F^{36, 734}$ statistic for the joint significance of the R&D variables in their natural form was a rather small 1.18 whereas that test statistic for the logarithmic form of the R&D variables was a highly significant 3.30. We, therefore, accepted the former and rejected the latter hypothesis and went on to test another simplification; whether or not the seven time dummies could be approximated by a linear time trend. The observed value of the $F^{30, 770}$ deviate for this hypothesis was .95, which is below the expected value of that test statistic given that the time dummies were in fact representing a simple trend. Two other hypotheses were tested but both were clearly rejected by the data. The first was that the distribution of the firm-specific constants was degenerate, that there were no "firm effects." After rejecting this hypothesis we went on to test whether it was reasonable to assume that the firm effects were uncorrelated with research expenditures. It was not. Thus the form of the equation we settled on was rather simple: the logarithm of patents as a function of a time trend, current and five consecutive lagged values of the logarithm of R&D expenditures, and (correlated) firm-specific constant terms.[2]

[2] There is one issue which we have not dealt with here because it is not very important in our sample. For observations where P = 0 log (P) is undefined. This exposes an underlying truncation problem in our model. That problem, however, is of minor importance for our sample since only 8 percent of the observations are at P = 0 and this is less than the percentage of observations at P = 1 (14%) indicating that the truncation problem is not large. It is even smaller for the larger R&D firm sample (N = 121) where the zero patents percentage is only 3. As a result we treated the whole problem as one of finding a point on the logarithmic scale for P = 0, and this was accomplished by adding a dummy variable to the independent variables for observations where P = 0. The estimated coefficients of this dummy variable are stable across models implying roughly the value of 0.1 to 0.7 for the P = 0 observations. It does raise the issue, though, whether our functional form (log-log) is appropriate for low patenting level observations. We intend to investigate more explicitly probabilistic models of the patenting process in subsequent work.

References

Box, G.E.P. and D.R. Cox, "An Analysis of Transformations," Journal of the Royal Statistical Society, Series B, 1962, 00. 211-243.

Comanor, W.C. and F.M. Scherer, "Patent Statistics as a Measure of Technical Change," Journal of Political Economy, Vol. 77, No. 3, May/June 1969, pp. 392-398.

Dictionary of Corporate Affiliation, National Register Publishing Company, Skokie, Illinois, 1972, 1976.

Griliches, Z., "R&D and the Productivity Slowdown," American Economic Review, Vol. 70, No. 2, May 1980, pp. 343-348.

___________, "Returns to Research and Development Expenditures in the Private Sector," in J.W. Kendrick and B.N. Vaccara (eds.) New Developments in Productivity Measurement and Analysis, National Bureau of Economic Research, University of Chicago Press, 1980, Chap. 8, pp. 419-454.

Kuznets, S., "Inventive Activity: Problems of Definition and Measurement," in R.R. Nelson (ed.) The Rate and Direction of Inventive Activity: Economic and Social Factors, National Bureau of Economic Research, Princeton University Press, 1962, pp. 19-51.

Mergers and Acquisitions, Journal of Corporate Venture, McLean, Virginia Vol. 8-11, 1974-1977.

Mundlak, Y., "On the Pooling of Time Series and Cross Section Data," Econometrica, Vol. 46, No. 1, January 1978, pp. 69-86

Mundlak, Y. and I. Hoch, "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," Econometrica, Vol. 33, No. 4, October 1965, pp. 829-841.

Nadiri, M.I. and G.C. Bitros, "Research and Development Expenditures and Labor Productivity at the Firm Level: A Dynamic Model," in J.W. Kendrick and B.N. Vaccara (eds.), New Developments in Productivity Measurement and Analysis, National Bureau of Economic Research, University of Chicago Press, 1980, pp. 387-412.

National Science Foundation, 1972, A Price Index for Deflation of Academic R&D Expenditures, NSF 72-310, Washington, D.C.

Nelson, R.R., (ed.), The Rate and Direction of Inventive Activity: Economic and Social Factors, National Bureau of Economic Research, Princeton, Princeton University Press, 1962.

Pakes, A., Economic Incentives in the Production and Adoption of Knowledge, Ph.D. Thesis, Harvard University, 1978.

Pakes, A. and M. Schankerman, "The Rate of Obsolescence of Knowledge, Research Gestation Lags, and the Private Rate of Return to Research Resources," Harvard Institute of Economic Research Discussion Paper No. 659, October 1978.

Rapoport, J., "The Anatomy of the Product-Innovation Process: Cost and Time," in E. Mansfield et al., Research and Innovation in the Modern Corporation, New York: Norton, 1971, pp. 110-135.

Scherer, F.M., "Firm Size, Market Structure, Opportunity, and the Output of Patented Innovations," American Economic Review, LV, No. 5, Part 1, 1965, pp. 1097-1125.

Standard and Poor Compustat, Industrial Compustat, 1978.

Taylor, C.T. and Z.A. Silberston, The Economic Impact of the Patent System, 1978, Cambridge University Press.

Temin, P., "Technology, Regulation, and Industrial Structure in the Modern Pharmaceutical Industry," Unpublished paper, MIT, 1979.

Wagner, L.U., :Problems in Estimating Research and Development Investment and Stock," American Statistical Association, Proceedings of the Business and Economic Statistics Section, 1968, pp. 189-198.