A Penny for Your Quotes: Patent Citations and the Value of Innovations

Author(s): Manuel Trajtenberg

Source: *The RAND Journal of Economics*, Spring, 1990, Vol. 21, No. 1 (Spring, 1990), pp. 172-187

Published by: Wiley on behalf of RAND Corporation

Stable URL: https://www.jstor.org/stable/2555502

RAND Journal of Economics
Vol. 21, No. 1, Spring 1990

A penny for your quotes: patent citations and the value of innovations

Manuel Trajtenberg*

The use of patents in economic research has been seriously hindered by the fact that patents vary enormously in their importance or value, and hence, simple patent counts cannot be informative about innovative output. The purpose of this article is to put forward patent counts weighted by citations *as indicators of the value of innovations, thereby overcoming the limitations of simple counts. The empirical analysis of a particular innovation (Computed Tomography scanners) indeed shows a close association between citation-based patent indices and independent measures of the social value of innovations in that field. Moreover, the weighting scheme appears to be nonlinear (increasing) in the number of citations, implying that the informational content of citations* rises *at the margin. As in previous studies, simple patent counts are found to be highly correlated with contemporaneous R&D; however, here the association is within a field over time rather than cross-sectional.*

1. Introduction

■ The study of technological change has been hampered all along by the scarcity of appropriate data and, in particular, by the lack of good indicators of innovation having a wide coverage. Patents would seem to be the one important exception, since they are the only manifestation of inventive activity covering virtually every field of innovation in most developed countries and over long periods of time. Yet, their use in economic research has not lived up to expectations primarily because patents exhibit an enormous variance in their "importance" or "value," and hence, simple patent counts cannot be very informative of innovative "output" (which is usually what we are after).

The goal here is to readdress this problem by examining the usefulness of patent indicators in the context of a particular innovation, Computed Tomography scanners, one of the most important advances in medical technology of recent times. The central hypothesis is that patent citations (i.e., references to patents appearing in the patent documents themselves), long presumed to be indicative of something like technological importance, may be informative of the economic value of innovations as well. Indeed, patent counts weighted by a citations-based index are found to be highly correlated (over time) with independent measures of the social gains from innovations in Computed Tomography. On the other hand, as in the previous literature, simple patent counts are found to be indicative only of the input side, as reflected in R&D outlays. Beyond establishing their role as indicators, the

* Tel-Aviv University and Harvard Business School.
I am grateful to Tim Bresnahan, Zvi Griliches, an anonymous referee, and to participants in the productivity workshop of the NBER Summer Institute for helpful comments on an earlier draft.

findings suggest that citations may reflect a sort of causal relationship between citing and cited patents, which is consistent with the view of innovation as a continuous and incremental process, punctuated by occasional breakthroughs.

The article is organized as follows. Section 2 lays out the basis for the use of patent citations. Section 3 offers a first view of the patent data in Computed Tomography, and Section 4 briefly describes the measures of the value of innovations in this field borrowed from a previous study (Trajtenberg, 1989). The tests of the main hypotheses are conducted in Section 5, and Section 6 offers some closing thoughts. Finally, in view of the fact that the number of citations per patent decreases drastically over time, I present in the Appendix a statistical procedure to test for "age versus importance" and for truncation effects.

2. Using patent counts and patent citations

■ It has long been thought that the detailed information contained in the patent documents may have a bearing on the importance of the innovations disclosed in them and that it may therefore be possible to construct patent indicators that could serve as proxies for the value of innovations.[1] Up to now, though, virtually the only patent measures used in economic research have been simple patent counts (henceforth SPC), that is, the number of patents assigned over a certain period of time to firms, industries, countries, *etc.*

The body of evidence that has accumulated since Schmookler (1966) indicates fairly clearly that SPC are closely associated with the input side of the innovative process, primarily with contemporaneous R&D expenditures in the cross-sectional dimension (Griliches, 1984). On the other hand, the few attempts to relate those counts to value indicators (e.g., the market value of innovating firms) have been largely unsuccessful. (See, for example, Griliches *et al.,* 1988.) As suggested above, those findings are hardly surprising, considering that patents vary enormously in their technological and economic significance. Thus, the mere counting of patents at any level of aggregation cannot possibly render good value indicators: simple patent counts assign a value of one to all patents by construction, whereas their true values exhibit a very large variance. Furthermore, there is substantial evidence to the effect that the distribution of patent values is highly skewed toward the low end, with a long and thin tail into the high-value side. As Scherer (1965) notes, those Pareto-like distributions might not have finite moments and, in particular, they might not have a finite variance; clearly, that would make the use of patent counts as proxies even more problematic. It is important to emphasize that those problems are inherent to the patent system as such,[2] and therefore definite solutions can hardly be expected.

An idea that has often been suggested is to use patent *citations* as an index of the importance or value of patents,[3] i.e., to count the number of times that each patent has been cited in subsequent patents and use the number to compute weighted patent counts. (I am referring to the citations appearing on the front page of patents under *References Cited.*) The potential significance of patent citations can be inferred from the following quotation:

During the examination process, the examiner searches the pertinent portion of the "classified" patent file. His purpose is to identify any prior disclosures of technology . . . which might anticipate the claimed invention and

[1] By value I mean the social benefits generated by the innovation in the form of the additional consumer surplus and the profits stemming from the innovation. The "value," "output," and "magnitude" of innovations are taken to mean exactly the same thing. (For a detailed discussion of those concepts, see Trajtenberg (1990).)

[2] This is so because the importance of a patent—however defined—can hardly be assessed *ex ante* and because it is not the task of patent examiners to make sure that the patents granted are of comparable worth.

[3] This idea draws from the extensive use of citations from and to scientific publications in the context of bibliometric studies. (See, for example, Price (1963).)

preclude the issuance of a patent; which might be similar to the claimed invention and limit the scope of patent protection . . . ; or which, generally, reveal the state of the technology to which the invention is directed . . . If such documents are found they are made known to the inventor and are "cited" in any patent which matures from the application. . . . Thus, the number of times a patent document is cited may be a measure of its technological significance. (Office of Technology Assessment and Forecast, 1976, p. 167)

Moreover, there is a legal dimension to patent citations, since they represent a limitation on the scope of the property rights established by a patent's claims, which carry weight in court. Equally important, the process of arriving at the final list of references, which involves the applicant and his attorney as well as the examiner, apparently does generate the right incentives to have all relevant patents cited, and only those. (See Campbell and Nieves (1979).) The presumption that citation counts are potentially informative of something like the technological importance of patents is thus well grounded.

The question is whether citations counts may also be indicative of the (*ex post*) value of the innovations disclosed in the cited patents.[4] This can only be answered empirically, but one can advance some further arguments that would strengthen the prior. Most patents cited are referenced in patents issued within the same narrowly defined field of innovation as the cited patents ("within citation"). The very existence of those later patents attests to the fact that the cited patents opened the way to a technologically successful line of innovation. Moreover, it presumably attests also to economic success (at least in expected value terms), since those subsequent patents are the result of costly innovational efforts undertaken mostly by profit-seeking agents. Given that citations to a patent are counted for a period of a few years following its issuance, there should be enough time for the uncertainty regarding the economic value of the innovation to resolve itself. Thus, if citations keep coming, it must be that the innovation originating in the cited patent had indeed proven to be valuable.

Whatever their merits, patent citations have rarely been used in economic research[5] probably because it used to be quite difficult to obtain the necessary data (i.e., the frequency of citations for each patent studied). Today, however, this is easily done with the aid of computerized search techniques, such as those offered by DIALOG or BRS. More important, the significance of a citations-based index can be ascertained only by relating it to independent measures of the value of innovations; given the scarcity of such measures, no firm conclusions could have possibly been drawn from citations-based statistical findings.[6] It is here that the advantage of having estimates of the social gains from innovation in Computed Tomography scanners from a previous study proved to be crucial. A further advantage is that both the patent counts and the value measures used to validate them refer in this case precisely to the same stretch of innovative activity, i.e., to advances in a carefully circumscribed product class and time period. Thus, the usual problems that arise when trying to match information belonging to disparate units (as often happens in this context) are altogether absent here.

Granted the use of citations, the next issue is how to go about constructing a sensible weighting scheme. A straightforward possibility is to weight each patent i by the actual number of citations that it subsequently received, denoted by C_i. Thus, if I were to compute

[4] This clearly need not be the same as technological importance: the latter could be thought of as having to do only with the supply side of innovations, whereas value obviously reflects a market equilibrium.

[5] An exception is Lieberman (1987). Other studies using patent citations in a related way (i.e., seeking to approximate something like importance) but not quite in the realm of economics proper include Carpenter *et al.* (1981), Ellis *et al.* (1978), and Narin and Wolf (1983).

[6] A series of articles in the press proclaimed some time ago that ". . . Japanese Outpace Americans in Innovation," on the basis of findings showing that Japanese patents have been cited more often than U.S. patents. (See *New York Times,* (March 7, 1988) and *Time,* (March 21, 1988); the study cited was conducted by Computer Horizons, Inc. for the National Science Foundation.) While that may be so, the mere finding of a higher frequency of citations does not and cannot prove anything by itself.

an index of weighted patent counts (WPC) for, say, a given product class in a given year, t, I would have,

$$WPC_t = \sum_{i=1}^{n_t} (1 + C_i), \tag{1}$$

where n_t is the number of patents issued during year t in that product class. This linear weighting scheme then assigns a value of one to all citations and all patents. Lacking more information on the citation process, this is a natural starting point but certainly not the only possibility; in fact, in Section 5, I compute a nonlinear index that allows for the existence of "returns to scale" in the informational content of citations.

3. Patents in Computed Tomography: a first look

■ Computed Tomography (CT) is a sophisticated diagnostic technology that produces cross-sectional images of the interior of the body, allowing the visualization of a wide range of organs with great accuracy. It has been hailed as one of the most remarkable medical innovations of recent times, comparable to the invention of radiography. Originally developed at the British firm EMI in the early seventies, CT soon attracted some twenty other firms worldwide, and the fierce competition that ensued brought about a breathtaking pace of technical advance. The diffusion of the new systems also proceeded very quickly: first introduced in the U.S. in 1973, by 1985 almost 60% of hospitals (community hospitals with more than 100 beds) had at least one system installed. The pace of innovation subsided in the late seventies as the technology matured and ceded its dominant place to new developments, particularly to Magnetic Resonance Imaging.

Searching in the PATDATA database (available through BRS), I located and retrieved all U.S. patents granted in Computed Tomography from the very start of the field in 1971[7] up to the end of 1986, totalling 456 patents.[8] I am quite certain that this is indeed the complete set and that it includes patents in CT only.[9] Clearly, having a clean set of patents, and hence clean patent-based indicators, is crucial in order to assess the usefulness of those indicators; otherwise, it would be impossible to tell whether the results are spurious (due to errors of measurement) or reflect real phenomena.

As is by now standard practice, patents are dated according to their application, rather than granting, date. Examining the distribution of lags between these two dates, I concluded that the patent data comprise virtually all patents applied for up to the end of 1982, 96% of the patents applied for in 1983, and smaller percentages of those applied for in later years. Thus, the analysis will be confined to the period of 1972–1982, although the citations appearing in the 1983–1986 patents will be taken into account as well.

Citation counts can be done in two ways: counting all citations or just those appearing in the set of patents belonging to the same field. In the "within referencing" case, the citation counts will be associated with the value of the innovations in the specific technological field to which they belong. On the other hand, an all-inclusive count will presumably capture

[7] In this case, it was easy to identify the very first patent: the origin of Computed Tomography is unequivocally associated with its invention by G. Hounsfield, as described in his U.S. patent #3778614, applied for in December 1971. Since there were no patents in CT in 1972, I shall treat this first patent as if it had been applied for in January 1972, rather than in December 1971, in order to avoid an unnecessary discontinuity in the data points.

[8] See Trajtenberg (1990) for a detailed discussion of the long-standing classification problem (i.e., how to match patents to economic categories) and for the advantages of using computerized search techniques in order to tackle it.

[9] The computerized search actually produced 501 patents, but 45 of them were eliminated after a careful examination of their abstracts; thus, I am certain that all the patents included do belong to CT. Furthermore, I am confident that the set includes all the relevant patents, since I was able to cross-check with other sources, including listings of patents from the manufacturers of CT scanners.

the value "spilled-over" to other areas as well. Given that the measures of innovation to be used in conjunction with the patent data refer to the gains from advances in CT as such, with no attempt to account for spillovers, I shall use the within referencing count.[10]

The first two columns of Table 1, graphically displayed in Figure 1, show the basic patent data to be used throughout. Note the smooth, cycle-like path followed by the yearly count of patents: it rises quite fast after 1973, peaks in 1977, and then declines steadily, carrying forward a thin tail. Notice also that the weighting scheme strongly influences the shape of the distribution, shifting it back toward the earlier period. In fact, the difference in the means of the two distributions is seventeen months; that is, the weighting scheme centers the action in mid-1976 rather than in late 1977. Given the very fast pace at which the CT technology evolved and that the period is just eleven years long, a difference of one and one-half years in the means is certainly very significant. Clearly, this shift is due to the fact that earlier patents were cited more frequently than later ones: as Table 1 shows, the average number of citations per patent went down from 72 to less than one, and the percentage of patents with no citations increased from zero to 92%.

The question is whether the observed citation frequencies are to be regarded as real phenomema, presumably reflecting something like the importance of patents, or just as statistical artifacts, induced by the mere passage of time. Two concerns arise in this context. First, it could be that older patents are cited more often simply because they have had more opportunities to be cited, since they precede a larger set of patents that could cite them. Second, given that CT is an ongoing technology (albeit already mature), it is quite certain that additional patents have been issued since the time of the search and that more will be granted in the future. Thus, the data set is necessarily truncated, which might bias downward the citation counts of recent patents.

TABLE 1 **Patents in Computed Tomography: Counts and Citations by Year**

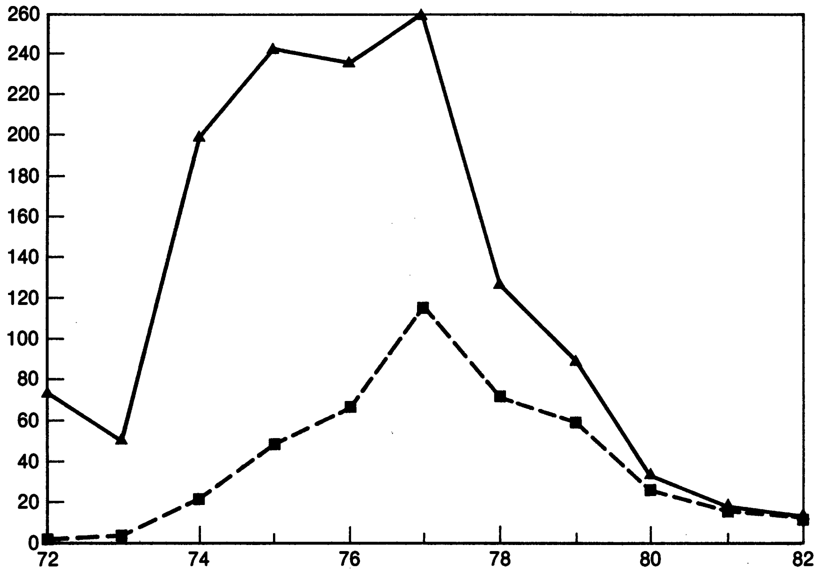
	Patents		Citations		
				Percentage of Patents with:	
Year	Simple Counts (*SPC*)	Weighted by Citations (*WPC*)	Average No. per Patent	0	5+
1972	1	73	72.0	0.0	100.0
1973	3	50	15.7	0.0	100.0
1974	21	199	8.5	4.8	76.2
1975	48	242	4.0	12.5	47.9
1976	66	235	2.6	21.2	22.7
1977	115	260	1.3	45.2	11.3
1978	71	126	0.8	54.9	4.2
1979	59	88	0.5	66.1	0.0
1980	26	33[a]	0.3	84.6	0.0
1981	15	18[a]	0.2	86.7	0.0
1982	12	13[a]	0.1	91.7	0.0
1983[b]	13	14	•	•	•
1984[b]	6	6	•	•	•
All	456	1357	2.1[c]	45.1[c]	16.2[c]

[a] These figures are slightly biased downwards. (See the Appendix).
[b] These are partial figures.
[c] These are averages.

[10] In this case, it would not have mattered much which count was used: in a sample of 30 patents in CT, the correlation between the two counts was found to be .99. Similarly, Campbell and Nieves (1979) report a correlation of .73 between what they called "in-set" and total citations for some 800 patents in the field of catalytic converters.

FIGURE 1

PATENTS IN CT: SIMPLE COUNTS, AND COUNTS WEIGHTED BY CITATIONS



■ Simple patent count (Source: Table 1, Column 1).

▲ Count weighted by citations (Source: Table 1, Column 2).

These are serious *a priori* objections that may arise whenever one tries to attach any meaning to citations data, and therefore deserve careful scrutiny.[11] The Appendix analyzes them in detail and shows that neither age nor truncation could possibly account for the observed distribution of citation counts. The issue of age is tackled by constructing a hypothetical "iso-important" distribution of citations and testing it against the observed distribution with the aid of a χ^2-test: the null hypothesis that older patents received more citations just because of the passage of time is rejected by a wide margin. As for the effect of truncation, the magnitude of the biases is estimated by extrapolating from the observed distribution of citation lags and application-granting lags. The main finding is that a bias does exist, but the absolute expected number of missing citations to recent patents is very small and hence could not possibly affect the results of the statistical analysis based on citation counts.

4. Estimates of the value of innovations in Computed Tomography

■ As stated earlier, I intend to relate patent counts weighted by citations to independent measures of innovation in CT taken from an earlier study (Trajtenberg, 1989). The basic idea behind those measures is as follows. Consider a technologically dynamic product class as it evolves over time, and assume that the different brands in it can be described in terms of a small number of attributes. Product innovation can then be thought of in terms of changes in the set of available products, both in that new brands appear and that the qualities of existing products improve. Applying discrete choice models to data on sales per brand

[11] The issue of age versus importance (closely related to Price's "immediacy factor") has commanded a great deal of attention in the scientometric literature. However, to the best of my knowledge, it has not yet been addressed with rigorous statistical tests. (See, for example, Line (1970) and Campbell and Nieves (1979).)

and on their attributes and prices, one can estimate the parameters of the demand functions and, under some restrictions, of the underlying utility function. The social value of the innovations occurring between two periods can then be calculated as the benefits of having the latest choice set rather than the previous one in terms of the ensuing increments in consumer and producer surplus. That is, given an estimated social surplus function, $W(\cdot)$, and the sets of products offered in two successive periods, S_t and S_{t-1}, the value of innovation would be measured by $\Delta W_t = W(S_t) - W(S_{t-1})$.

The model used to estimate the function $W(\cdot)$ was the multinomial logit model, rendering the well-known choice probabilities $\pi_j = \exp\phi(z_{jt}, p_{jt}) / \sum_i^m \exp[\phi(z_{it}, p_{it})]$, where z is the vector of attributes, p is price, m is the number of alternatives in the choice set, and $\phi(\cdot)$ is the branch of the indirect utility function related to the product class in question. Integrating those probabilistic demand functions, one obtains measures of consumer surplus of the form[12]

$$W_t = \ln \{ \sum_i^{m_t} \exp[\phi(z_{it}, p_{it})] \} / \lambda_t, \tag{2}$$

where λ stands for the marginal utility of income. The differences ΔW_t are then computed from (2) for every pair of adjacent years. Noting that ΔW_t refers to the gains accruing to the representative consumer, I also computed the total gains associated with the innovations at t, using

$$TW_t = \Delta W_t \left[n_t + K(\sum_{\tau=0}^{t} \Delta W_\tau) \int_{t+1}^{\infty} f(\tau) e^{-r(\tau-t-1)} d\tau \right] \equiv \Delta W_t(n_t + n_f), \tag{3}$$

where n_t is the number of consumers at t, $K(\cdot)$ is the ceiling of the diffusion curve (that shifts up as a consequence of successive innovations), $f(\cdot)$ is the diffusion path, and r is the interest rate. Thus, TW multiplies ΔW_t by the current and future number of consumers that benefit from the innovations at t, the latter being assessed on the basis of the observed diffusion process.

For the earlier study, I gathered a comprehensive data set on CT scanners, including the prices and attributes of all scanners marketed in the U.S. since the inception of CT in 1973 up to 1982, and details of all sales to U.S. hospitals (over 2,000 observations). Applying the methodology just sketched to these data, I obtained yearly estimates of ΔW_t and of TW_t, as shown in Table 2 along with other data on CT. Note that ΔW_t and TW_t are very large at first and then decline sharply, carrying forward a thin tail. Thus, the bulk of gains from innovation were generated during the earlier years of the technology even though the action in terms of R&D and entry peaks later on.

5. Patents as indicators of innovation: the statistical evidence

■ The question to be addressed here is, then, To what extent can patent-based indices (denoted by P) serve as indicators of the value of innovations as measured by ΔW (or

[12] In the present case, ΔW was confined to changes in consumer surplus, since the net aggregate profits of CT manufacturers were actually nil over the period studied. This may raise questions as to the rationale of the expected link between ΔW and patent indices, since, from the point of view of the incentives to innovate (and hence, to patent), what counts are the benefits appropriated by the firms, not consumer surplus. This is true *ex ante*; that is, firms innovate because they expect to be able to appropriate a great deal of the social surplus, but competition may significantly reduce (*ex post*) that portion. (Recall that competition in the CT market was indeed very intense during the period studied.) In that case one would still expect to find a significant correlation between total surplus and, say, *WPC*, even if the former is made just of consumer surplus.

TABLE 2 **Measures of Innovation and Other Data on Computed Tomography Scanners**

Year	ΔW	TW	*R&D*	Number of Firms	Number of New Brands	Number of New Adopters
1973	2.99	638	20.6[a]	3	1	16
1974	8.71	6926	22.6	8	1	74
1975	1.51	1503	59.7	12	4	216
1976	4.78	5959	96.1	13	11	317
1977	0.94	997	79.7	14	14	328
1978	0.12	79	64.3	11	6	211
1979	0.14	73	56.1	9	5	209
1980	0.07	30	46.4	8	2	177
1981	0.18	79	37.9	8	3	101
1982	0.20	87	37.9	8	8	•

Source: Trajtenberg (1989).
The figures for ΔW, TW, and *R&D* are in millions of constant 1982 dollars.
[a] This figure refers to total R&D expenditures from 1968 through 1973.

TW).[13] By indicators I formally mean predictors, and hence, the statistical criterion to be used is the mean-squared prediction error: $MSE = E(\Delta W - P)^2$. Assuming that the bivariate distribution $f(\Delta W, P)$ is such that the regression function of ΔW on P is linear (e.g., a bivariate normal), then the MSE of the (best) linear predictor is just $\sigma^2_{\Delta W}(1 - \rho^2)$, where ρ^2 is the correlation coefficient between ΔW and P. (See Lindgren (1976).) Thus, in probing the adequacy of alternative patent indicators, I used the correlation coefficient as the sole criterion.[14] The hypotheses tested are (note that they refer to a given technological field, or product class, as it evolves over time) as follows:

Hypothesis 1. Patent counts weighted by citations (WPC) are good indicators of the value of innovations as measured by ΔW or TW, but simple counts (SPC) are not.

Hypothesis 2. Simple patent counts are good indicators of the inputs to the innovative process as measured by R&D expenditures.

In order to examine them, I considered two alternative patent counts: one including all patents in CT and another based on patents granted just to CT manufacturers. (The latter accounted for 66% of all patents and 80% of all citations.) Since ΔW and TW were computed on the basis of the CT scanners actually marketed in the U.S., we would expect those measures to be more highly correlated with the patents granted just to firms in CT.[15] I also considered various lags between patent counts and the measures ΔW and TW. At issue here is the expected time interval between the application date of a patent and the appearance in the market of the innovation disclosed in that patent.[16] Since the interest of inventors is to file for a patent as early as possible, the actual timing must be effectively constrained by the stringency of the applications requirements set by the Patent Office.

[13] Recalling that ΔW refers to the incremental gains, whereas TW stands for total gains, it is not clear *a priori* which of them is more relevant in the present context.

[14] Less formally, since the maintained hypothesis is that the measures $\{\Delta W, TW\}$ accurately capture the value of innovations, the only remaining question is whether patents (which could be at best just an indirect manifestation of the same) closely follow the path of those variables over time. More pragmatically, with the small number of observations available, it would have been very hard to estimate anything but simple correlations.

[15] That would not be so only if the appropriability of the patents issued to the other assignees had been extremely low, i.e., only if CT manufacturers benefitted from the innovations done by other inventors as much as they did from their own.

[16] For a detailed discussion of this issue, see Trajtenberg (1990).

Those requirements are significantly more severe in the U.S. than in Europe; in fact, 56% of all patents in CT were applied for in various European countries (primarily the U.K. and Germany) before being filed in the U.S., with an average lag between the two applications of thirteen months. Thus, we would expect the lag between patent counts and ΔW to be quite short if patents are dated according to the application date in the U.S. (as done here), and over a year longer than that (on average) if the application date abroad is taken into consideration.

The data used in testing the hypotheses consist of yearly observations on ΔW, TW, and patent counts, covering the period of 1973–1982. (See Tables 1 and 2.) As mentioned above, this is when the bulk of the innovative action in CT took place; in fact, the pace of advancement had already subsided in the late seventies, and the technology barely changed during the eighties. Thus, the fact that the period examined here is only a decade long does not impair the validity of the statistical analysis performed: the history of innovation in CT is short, and hence, there are few observations; this is, however, essentially the whole history.

□ **Testing the first hypothesis.** Table 3 presents the correlations between patent counts and $\{\Delta W, TW\}$, with the former variables lagged between zero and six months.[17] The first and most important finding is that WPC_t is, in effect, correlated with the value measures of innovation, whereas SPC_t clearly is not, in all the cases considered. Thus, the evidence

TABLE 3 **Correlations[a] of Simple and Weighted Patent Counts with ΔW, TW**

	All Patents		Patents to Firms in CT	
Lags	ΔW	TW	ΔW	TW
	(a) With Weighted Counts			
Contemporary	0.509 (0.13)	0.587 (0.07)	0.616 (0.06)	0.626 (0.05)
3 Months	0.513 (0.13)	0.635 (0.05)	**0.685** (0.03)	**0.755** (0.01)
4 Months	0.480 (0.16)	0.600 (0.07)	0.677 (0.03)	0.744 (0.01)
6 Months	0.317 (0.37)	0.466 (0.17)	0.495 (0.15)	0.605 (0.006)
	(b) With Simple Counts			
Contemporary	−0.162 (0.65)	0.032 (0.93)	−0.087 (0.81)	0.093 (0.80)
3 Months	−0.198 (0.58)	0.006 (0.99)	−0.076 (0.83)	0.131 (0.72)
6 Months	−0.283 (0.43)	−0.090 (0.81)	−0.175 (0.63)	0.027 (0.94)

[a] Pearson correlation coefficients.
Significance levels for H_0: corr = 0 are given in parentheses.

[17] Even though the figures are annual, the lag could be varied by monthly increments, since the patent data are virtually continuous over time. Note also that since the ΔW series begins in 1973, I just added the 1972 (first) patent to the patent count of 1973; that is, the ΔW figure for 1973 refers to the first CT scanner marketed, and hence, it obviously corresponds to the initial patents in the field, including the very first.

provides ample support for the first hypothesis. Second, the correlations increase substantially as I narrow the scope of the (weighted) counts to the patents granted to firms in CT. This implies, as suggested, that the appropriability of patents awarded to other assignees was not nil. Third, the correlations peak when the patent counts are lagged just one quarter, declining monotonically as the lag increases.[18] Superimposing the mean foreign-U.S. application lag of thirteen months mentioned earlier, one obtains an overall lead time of sixteen months. This is consistent with the intense technological rivalry that characterized the market for CT scanners in the seventies.

Returning to the basic finding of a high correlation between WPC_t and ΔW_t, I can now (re)interpret the distribution of citation counts across patents as an implied distribution of the value of innovations. As shown in Table 4, the observed distribution fits well the received wisdom on this matter (see, for example, Pakes and Schankerman (1984) and Pakes (1986)): it is very skewed, with almost half the patents never cited (and hence, of little *ex post* value) and a lucky few being worth a great deal.[19] Thus, contrary to the often-voiced view that patent data cannot possibly capture important innovations, the results here show that citation counts can span well the whole range of innovations.[20]

TABLE 4 Distribution of Patents According to Number of Citations

Number of Citations	Number of Patents	Percentage of Patents	Cumulative Percentage
0	215	47.1	47.1
1	78	17.1	64.3
2	54	11.8	76.1
3	35	7.7	83.8
4	21	4.6	88.4
5	10	2.2	90.6
6	15	3.3	93.9
7	8	1.8	95.6
8	3	0.7	96.3
9	3	0.7	96.9
10	2	0.4	97.4
12	1	0.2	97.6
13	2	0.4	98.0
14	1	0.2	98.2
16	1	0.2	98.5
17	2	0.4	98.9
19	1	0.2	99.1
20	1	0.2	99.3
21	1	0.2	99.6
25	1	0.2	99.8
72	1	0.2	100.0

[18] Recall, however, from footnote 17 that the 1972 patent was simply added to the 1973 patents in computing the correlations. Thus, the first lag was actually longer (about one year long), and the overall lag would increase from three to four months if one averages that first lag with the rest.

[19] Campbell and Nieves (1979) present the distribution of citations for all U.S. patents issued from 1971 to 1978, and Narin (1983) does the same for 13,264 chemical and allied product patents issued in 1975. Both distributions look remarkably similar to the one for CT scanners. Unfortunately, the citation values in Campbell and Nieves (1979) only go up to 13+, and therefore, I cannot be sure whether the distribution for CT is typical in its upper tail.

[20] According to Scherer (1965), ". . . patent statistics are likely to measure run-of-the-mill industrial inventive output much more accurately than they reflect the occasional strategic inventions which open up new markets and new technologies." Of course, Scherer was quite right at the time, given the kind of data available then.

□ **Nonlinearities in citations.** In constructing the weighting index, I have implicitly assumed so far that a citation is worth as much as a patent, i.e., that the weights are linear in the number of citations. However, there may be something akin to increasing or decreasing returns to the informational content of citations, in which case the weighting scheme would be nonlinear. Consider the more general specification

$$WPC_t(\alpha) = \sum_{i=1}^{n_t} (C_i^\alpha + 1) = n_t + \sum_{i=1}^{n_t} C_i^\alpha, \qquad \alpha > 0.$$

Notice that the nonlinearity is assumed to occur in the citations to each patent, rather than in some patent aggregate (e.g., in the yearly counts), since the results would otherwise be highly sensitive to the chosen aggregation scheme, which is rather arbitrary. The problem now is to find α_1^* and α_2^* such that [21]

$$\alpha_1^* = \underset{\alpha}{\operatorname{argmax}} \operatorname{corr} [WPC_t(\alpha), \Delta W_t] \quad \text{and} \quad \alpha_2^* = \underset{\alpha}{\operatorname{argmax}} \operatorname{corr} [WPC_t(\alpha), TW_t].$$

As can be seen in Table 5, the maxima occur when $\alpha_1^* = 1.3$ and $\alpha_2^* = 1.1$; that is, the weighting scheme is in fact convex in the number of citations per patent. (Note, however, that the relationship between WPC_t and ΔW_t is still linear.) This finding means that the marginal informational content of WPC_t increases with the number of citations, strengthening the potential role of WPC_t as an indicator of the value of innovations. Furthermore, it implies that the variance in the value of patents is larger and that the distribution of those values is more skewed than what could be inferred from the simple count of citations.

TABLE 5 **Correlations[a] of *WPC* with ΔW and *TW*: Searching for Nonlinearities**

Exponent	Patents to Firms in Computed Tomography				All Patents	
	Contemporary		Lagged Three Months		Lagged Three Months	
α	ΔW	TW	ΔW	TW	ΔW	TW
0.80	0.455	0.543	0.512	0.653	0.329	0.503
	(0.19)	(0.10)	(0.13)	(0.04)	(0.35)	(0.14)
0.90	0.538	0.590	0.601	0.711	0.419	0.570
	(0.11)	(0.07)	(0.07)	(0.02)	(0.23)	(0.09)
1.00	0.616	0.626	0.685	0.755	0.513	0.635
	(0.06)	(0.05)	(0.03)	(0.01)	(0.13)	(0.05)
1.10	0.680	0.642	0.754	0.777	0.605	0.687
	(0.03)	(0.05)	(0.01)	(0.008)	(0.06)	(0.03)
1.20	0.721	0.635	0.798	0.770	0.684	0.719
	(0.02)	(0.05)	(0.006)	(0.009)	(0.03)	(0.02)
1.30	**0.738**	0.607	**0.813**	0.736	0.738	**0.720**
	(0.01)	(0.06)	(0.004)	(0.02)	(0.02)	(0.02)
1.40	0.730	0.560	0.800	0.677	**0.760**	0.689
	(0.02)	(0.09)	(0.006)	(0.03)	(0.01)	(0.03)
1.50	0.703	0.501	0.766	0.606	0.751	0.634
	(0.02)	(0.14)	(0.01)	(0.06)	(0.01)	(0.05)
1.60	0.663	0.436	0.718	0.527	0.719	0.562
	(0.04)	(0.21)	(0.02)	(0.11)	(0.02)	(0.09)

[a] Pearson correlation coefficients.
Significance levels for H_0: corr = 0 are given in parentheses.

[21] Notice that it is not possible to estimate those exponents by, say, running a regression in the logs, since the nonlinearity refers to the citations to each patent.

Notice that the results that $\alpha_i > 1$, $i = 1, 2$, are robust (that is, they occur also when there is no lag and when all patents, rather than patents to firms in CT, are used), and that α_1^* and α_2^* are global maxima. Note also that $\alpha_1^* > \alpha_2^*$; that is, the nonlinearity is stronger when patent counts are related to ΔW rather than to TW. Recall that ΔW is a measure of the gains to the representative consumer, whereas TW multiplies ΔW by the number of consumers that benefit from the innovation at present and in the future. Thus, the fact that $\alpha_1^* > \alpha_2^*$ and that corr $[WPC(\alpha_1^*), \Delta W] >$ corr $[WPC(\alpha_2^*), TW]$, can be taken to mean that citations are more informative of the value of innovations *per se,* rather than of the size of the market for the products embedding those innovations. This is reassuring, since we expect that the factors related to the technology itself (rather than to market size) will be dominant in the citing process.

□ **The second hypothesis.** The relationship between patents and R&D has been intensively scrutinized in past research,[22] and the results appear to be quite uniform, centering around the following stylized facts: (a) there is a strong statistical association between patents and R&D expenditures; (b) this relationship appears to be mostly contemporaneous; and (c) R&D explains a great deal of the cross-sectional variance in patenting but not much of the variation over time. The second hypothesis also postulates a close association between patents and expenditures on R&D, but within a given field over time rather than across firms or industries. From Table 6 we see first that there is indeed a high correlation between SPC_t and $R\&D_t$ and a much weaker one between $R\&D_t$ and WPC_t; the second hypothesis is therefore amply confirmed. Second, the degree of association peaks when $R\&D_t$ is lagged just five months, supporting previous findings of short gestation lags. Third, the correlations are slightly higher for counts of all patents than for patents to firms in CT, suggesting some degree of spillovers from the R&D done by CT manufacturers to other assignees.

It is also worth reporting the following correlations, which indicate that SPC_t tends to move together, not just with R&D, but also with other manifestations of the innovative

TABLE 6 **Correlations[a] Between Patent Counts and R & D**

	All Patents		Patents to Firms in Computed Tomography	
Lags	SPC_t	WPC_t	SPC_t	WPC_t
None	0.869 (0.0002)	0.609 (0.05)	0.843 (0.001)	0.525 (0.097)
Three months	0.919 (0.0001)	0.591 (0.04)	0.912 (0.0001)	0.495 (0.102)
Four months	0.924 (0.0001)	0.582 (0.05)	0.914 (0.0001)	0.491 (0.105)
Five months	**0.933** (0.0001)	0.577 (0.05)	**0.918** (0.0001)	0.483 (0.112)
Six months	0.921 (0.0001)	0.543 (0.07)	0.903 (0.0001)	0.450 (0.142)
One year	0.831 (0.0008)	0.248 (0.44)	0.794 (0.002)	0.152 (0.638)

[a] Pearson correlation coefficients.
Significance levels for H_0: corr = 0 are given in parentheses.

[22] Many of the articles in Griliches (1984) have to do with this issue; extensive references to previous works can also be found there.

action taking place in a given field over time. (All are contemporaneous; the data are taken from Table 2.)

$$\text{corr}\,(SPC_t, \text{Number of Firms in the CT Market}) = \underset{(.0007)}{.858}$$

$$\text{corr}\,(SPC_t, \text{Number of New Scanners Introduced in the Market}) = \underset{(.002)}{.813}$$

$$\text{corr}\,(SPC_t, \text{Number of New Adopters of CT}) = \underset{(.0002)}{.913.}$$

The first two correlations reflect the fact that competition in the CT market was driven primarily by rivalry in innovation, whereas the third has to do with the interaction between innovation and diffusion.

6. The usefulness and meaning of patent data: concluding remarks

■ The findings presented above suggest that patent citations may be indicative of the value of innovations and, if so, that they may hold the key to unlock the wealth of information contained in patent data. In order to understand the various roles that patent-based indicators may thus come to play, it may be helpful to use a familiar analogy, namely, to think of patents as working papers in economics and, accordingly, of economic departments as firms, of fields in economics as industries, and so forth. Working papers are "produced" roughly in proportion to the number of faculty, as patents are with respect to R&D. The fact is that it does not take much to get a patent once the firm has an established R&D facility going, as it does not take much to write a working paper. Still, a larger number of patents presumably indicates that much research efforts have been invested by the R&D staff, as more papers would suggest that the faculty is "trying harder." Thus, a simple patent count could be regarded as a more refined input measure (vis-a-vis R&D), in the sense that it incorporates part of the differences in effort and nets out the influence of luck in the first round of the innovative process. Of course, as with patents, a mere count of working papers written does not say much about the value of the scientific contributions made; for that, one would need information on whether and where they get published, the number of citations that they receive over time, etc. Clearly, those indicators would be to working papers what patent citations are to patents.

Beyond establishing the role of citations as indicators in a purely statistical sense, the results of this article can also be seen as lending support to a particular view of the innovative process, in which context citations are associated with real phenomena rather than just being a useful data contrivance. This view sees innovation as a continuous time process that has a predominantly incremental nature, punctuated by occasional breakthroughs that bring forth subsequent innovative efforts and direct them into novel channels.[23] Congruent with such a view, a patent would be regarded as important if it opened the way to a successful line of further innovations; the patents coming in its wake would naturally cite it, and hence, those citations could be taken as first-hand evidence of the path-breaking nature of the original patent. A particular case is when the important patent(s) refers to a new product (as, say, the first few patents by Hounsfield in CT): truly novel ideas are almost by definition crude and lacking at first but get better over time as a result of further research efforts. These efforts would generate down-the-line patents aimed at refining and improving the original

[23] This view stands in opposition to the innovative process being conceived of as a sequence of discrete, well-compartmentalized, and sizeable events that occur essentially in a random fashion from the point of view of technology itself. Such a view presupposes a one-to-one correspondence between each of those singular innovations and patents, leading to an interpretation of patent importance and of patent citations quite different from the one held here.

innovation and, again, these patents would be most likely to cite the basic one. The key point is that in this context, citing patents would bear a sort of causal relationship to the cited patent, with citations being the overt manifestation (instituted as common practice by the Patent Office) of such a link.

Once their meaning has been well established, the use of patent data may offer additional advantages in itself and over alternative data sources. First, patent data can be easily obtained all the way to the very beginning of a product class, whereas the gathering of conventional industry data usually starts only when a sector is well established. Thus, patent counts and citations may play an important role in studying the very emergence of new products, which seems to be the period when most of the important innovations occur. Second, patent data are richer, finer, and have a wider coverage than say, R&D expenditures, and are practically continuous in time.

All of my conclusions have been expressed in a qualified manner, since they are based upon the findings from a single case study. It is important to emphasize, however, that the sort of validation of the citation-based patent index attempted here could hardly have been done in a wider context simply because the measures of the value of innovations that would be required for that purpose are nowhere to be found. (If such measures were widely available, we would hardly need the more imperfect patent indicators.) I hope that future research along similar lines will bring in more supportive evidence and further demonstrate the attractiveness of the proposed indicators.

Appendix

■ A statistical analysis of truncation and age effects follows.

Testing age versus importance. The starting point for the test is the specification of a hypothetical citation process under which all patents are of equal importance, and hence the only differentiating factor is age. The distribution of citations thus generated could then be compared with the actual one, and the maintained hypothesis that all patents are equally important could be tested with the aid of a Pearson semiparametric χ^2-test. (See DeGroot (1975).) As a first step, patents are ordered according to their application date and indexed with $i = 1, \ldots, N$ ($N = 456$). (Thus, i indicates the cumulative number of patents in CT applied for up to patent i.) Denoting the probability that patent i will be cited in patent j by p_{ij} (for $i < j$) and the number of references to previous patents in CT appearing in patent j by r_j, I define the patents $1 \leq i < j$ to be iso-important if

$$p_{ij} = \frac{r_j}{j-1} = p_j, \qquad j = 1, \ldots, N. \tag{A1}$$

Thus, equal importance is taken to mean that all patents applied for up to a certain point in time have the same probability of being cited by a subsequent patent. In other words, (A1) means that the citations appearing in patent j are the result of r_j random drawings (without replacement) from a pool containing the j-1 patents that preceeded it.[24] Noting that (A1) also implies time independence (that is, for any $i < j < k$, p_{ik} is independent of p_{ij}), the expected number of citations of patent i can be computed simply as $C_i^e \equiv E(C_i) = \sum_{j=i+1}^{N} p_j$. Obviously, $C_i^e > C_j^e$ for any $i < j$; that is, older patents will get more citations on average than recent ones, just by virtue of their age. Notice also that p_j has to decrease eventually with j,[25] thus reinforcing the pure age effect. That is, not only do later patents miss the earlier p_j's, but those probabilities tend to be the large ones, a fact that further reduces the expected number of citations of recent patents vis-a-vis older ones.

In order to perform the χ^2-test, the data were aggregated by months, since it would be unreasonable to attach any significance (in the sense of differences in C_i^e) to the precise day of application. Indexing by τ and t the number

[24] Clearly, this is not the only possible definition of iso-importance. Notice, however, that by defining p_{ij} to be independent of the distance $(j - i)$, I implicitly favor the earlier patents, thus increasing the power of the test. That is, any plausible departure from (A1) would have the probabilities decrease with $(j - i)$, making the distribution of expected citations more uniform, and hence, making it easier to reject the null hypothesis.

[25] This must be true unless r_j increases indefinitely over time, which is highly unlikely; in the case of CT, r_j was quite stable over the whole period.

of months elapsed since January 1972, the observed number of citations is $C_t^o = \sum_{i=1}^{n_t} C_i$, where n_t is the number of patents in month t. Similarly, redefining (A1) in monthly terms gives

$$p_\tau = \sum_{i=1}^{n_\tau} r_i / \sum_{j=1}^{\tau-1} n_j, \tag{A2}$$

and so, $C_t^e = n_t \sum_{\tau=t+1}^{T} p_\tau$. Turning now to the test,[26]

$$\chi^2 = \sum_{t=1}^{156} \frac{(C_t^e - C_t^o)^2}{C_t^e} = 1025 >> 147 = \chi^2(110), \qquad \text{where} \qquad \alpha = .01.$$

Thus, the hypothesis that the observed distribution of citations is due solely to age is strongly rejected. As is to be expected, the largest discrepancies between actual and expected values occur at the very beginning of the period. In particular, the values for the first patent are $C_1^o = 72$, $C_1^e = 5.96$, and hence, $(C_1^e - C_1^o)^2/C_1^e = 731$, which amounts to .75 of the computed χ^2-statistic. Since this first patent can be regarded in many ways as an exception, the test was redone after deleting it; again, the null hypothesis is rejected by a wide margin.

Assessing the truncation bias. The other potential problem in this context is that the (unavoidable) truncation of the data might induce a bias in the citation counts; the extent of such bias, in turn, will depend upon the behavior of citation lags and the rate of new patent arrivals after the date of search. Citation lags refer to the length of time elapsed between the dates of the citing patents and of the cited patent: the shorter they are, the less severe the problem will be.[27] Denote the frequency distribution of citation lags by f_τ; for example, if year t patents are to receive (on average) C_t citations per patent, f_τ stands for the percentage of those citations to be received after τ years. (Thus, $\sum_{\tau=t}^{\infty} f_\tau = 1$.) Likewise, define $c_{t\tau} = f_\tau C_t$ and $g_{t\tau} = c_{t\tau}/n_\tau$, where n_τ is the total number of patents in year τ. Now, suppose that because of truncation, one can actually obtain only a fraction, h_τ, of them; then, assuming that $g_{t\tau}$ is invariant with respect to h_τ (i.e., that citations of year t patents are randomly distributed among the n_τ patents), the observed average number of citations to year t patents will be $c_{t\tau}^o = g_{t\tau} h_\tau n_\tau = h_\tau f_\tau C_t$. Thus, given the sequences $\{h_\tau, f_\tau\}$, one can compute for each year the fraction $v_t = \sum_{\tau=t}^{\infty} h_\tau f_\tau$, where v_t stands for the percentage of citations that patents in year t can be expected to receive out of the total that they would have received had it not been for the truncation of the data. Using the granting-application lags to obtain h_τ and the citation lags for f_τ, the expected biases are easily computed and presented in Table A1.

Thus, we do miss a few citations because of the truncation of the data; moreover, there is, as expected, a truncation bias in the sense that we have a smaller fraction of the true number of citations of later patents than of earlier ones. However, the absolute expected number of missing citations is very small, and hence, it is clear that

TABLE A1 Expected Biases

		Number of Citations		
Year of Cited Patents	v_t	Actual	Missing (Rounded)	Fraction Missing
Up to 75	1.000	491	0	0.00
76	0.998	169	0	0.00
77	0.990	145	1	0.01
78	0.969	55	2	0.04
79	0.930	29	2	0.07
80	0.861	7	1	0.14
81	0.732	3	1	0.33
82	0.527	1	1	1.00

[26] The summation is done in principle over 156 months, covering the thirteen years from 1972 to 1984. However, patents were actually applied for only in 111 months out of the 156, and hence, there are just 110 degrees of freedom (d). When d exceeds 100 (as it does here), the critical χ^2 value is to be computed as follows (see Harnett (1975)): $\chi^2 = 1/2(z_\alpha + \sqrt{2d-1})^2$, where z is the standardized normal deviate (for $\alpha = .01$, $z_\alpha = 2.3263$).

[27] In the present case, these lags are relatively short (the mean lag is three years), suggesting that the truncation problem is not too severe on that account. For a detailed discussion, see Trajtenberg (1990).

the truncation problem cannot possibly affect the conclusions of this article. (This holds true even if the bias had been for some reason underestimated by, say, a factor of two.)

References

CAMPBELL, R.S. AND NIEVES, A.L. *Technology Indicators Based on Patent Data: The Case of Catalytic Converters—Phase I Report: Design and Demonstration.* Battelle, Pacific Northwest Laboratories, 1979.

CARPENTER, M.P., NARIN, F., AND WOLF, P. "Citation Rates to Technologically Important Patents." *World Patent Information,* Vol. 3 (1981), pp. 160–163.

DEGROOT, M.H. *Probability and Statistics.* Reading, Mass.: Addison-Wesley, 1975.

ELLIS, P., HEPBURN, G., AND OPPENHEIM, C. "Studies on Patent Citation Networks." *Journal of Documentation,* Vol. 34 (1978), pp. 12–20.

GRILICHES, Z., ed., *R&D, Patents, and Productivity.* Chicago: University of Chicago Press, 1984.

———, HALL, B.H., AND PAKES, A. "R&D, Patents and Market Value Revisited: Is There a Second (Technological Opportunity) Factor?" Working Paper No. 2624, National Bureau of Economic Research, 1988.

HARNETT, D.L. *Introduction to Statistical Methods.* 2nd ed. Reading, Mass.: Addison-Wesley, 1975.

LIEBERMAN, M.B. "Patents, Learning-by-Doing, and Market Structure in the Chemical Processing Industries." *International Journal of Industrial Organization,* Vol. 5 (1987), pp. 257–276.

LINDGREN, B.W. *Statistical Theory.* New York: MacMillan, 1976.

LINE, M.B. "The Half-Life of Periodical Literature: Apparent and Real Obsolescence." *Journal of Documentation,* Vol. 26 (1970), pp. 46–52.

NARIN, F. "Patent Citation Analysis as an Analytical Tool in Marketing Research." CHI Research/Computer Horizons, Inc., 1983.

NARIN, F. AND WOLF, P. "Technological Performance Assessments Based on Patents and Patent Citations." CHI Research/Computer Horizons, Inc., 1983.

OFFICE OF TECHNOLOGY ASSESSMENT AND FORECAST, U.S. DEPARTMENT OF COMMERCE, PATENT, AND TRADEMARK OFFICE. *Sixth Report.* Washington D.C.: U.S. Government Printing Office, 1976.

PAKES, A. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica,* Vol. 55 (1986), pp. 755–784.

——— AND SCHANKERMAN, M. "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources." In Z. Griliches, ed., *R&D, Patents, and Productivity.* Chicago: University of Chicago Press, 1984.

PRICE, D.J. DE SOLLA. *Little Science, Big Science.* New Haven: Yale University Press, 1963.

SCHERER, F. "Using Linked Patent and R&D Data to Measure Interindustry Technology Flows." In Z. Griliches, ed., *R&D, Patents, and Productivity.* Chicago: University of Chicago Press, 1984.

———. "Firm Size, Market Structure, Opportunity, and the Output of Patented Innovations." *American Economic Review,* Vol. 55 (1965), pp. 1097–1123.

SCHMOOKLER, J. *Invention and Economic Growth.* Cambridge, Mass.: Harvard University Press, 1966.

TRAJTENBERG, M. *Economic Analysis of Product Innovation—The Case of CT Scanners.* Cambridge, Mass.: Harvard University Press, 1990.

———. "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners." *Journal of Political Economy,* Vol. 97 (1989), pp. 444–479.