

Discussion Paper No. 97-27

**Citation Frequency and the Value of
Patented Innovation**

Dietmar Harhoff, Francis Narin,
Frederic M. Scherer, Katrin Vopel

Non-technical Summary

Through a survey, economic value estimates were obtained on 962 inventions made in the United States and Germany and on which German patent renewal fees were paid to full-term expiration in 1995. A search of subsequent U.S. and German patents yielded a count of citations to those patents. Patents renewed to full term were significantly more valuable than patents allowed to expire before their full term. The higher an invention's economic value estimate was, the more the relevant patent was subsequently cited. Results from two wide-ranging surveys, one in Germany and one in the United States, support Trajtenberg's conclusion that patents with greater economic value are more heavily cited in subsequent patents. Our evidence suggests at least a two-stage escalation of economic values and citation counts. First, patents that are renewed to full term expiration in environments such as in Germany with highly progressive annual maintenance fees are more valuable and more highly cited than run-of-the-mill patents allowed to expire before running to full term. Second, within the relatively exclusive cohort of full-term patents, citation frequency rises with economic value, although with considerable noise in the relationship. The most cited patents turn out to be very valuable indeed, with a single U.S. citation implying on average more than \$1 million of economic value. These findings provide new support for research seeking to overcome the limitations of simple patent counts as a measure of innovative contribution by acquiring data on citation frequencies and the number of nations in which patent protection is sought.

Citation Frequency and the Value of Patented Innovation

by

Dietmar Harhoff, Francis Narin,
Frederic M. Scherer, Katrin Vopel*

August 1997

Abstract

Through a survey, economic value estimates were obtained on 962 inventions made in the United States and Germany and on which German patent renewal fees were paid to full-term expiration in 1995. A search of subsequent U.S. and German patents yielded a count of citations to those patents. Patents renewed to full term were significantly more valuable than patents allowed to expire before their full term. The higher an invention's economic value estimate was, the more the relevant patent was subsequently cited.

Keywords

Innovation, Patent, Citation, Count data model

JEL classifications

O34 and O47.

* Centre for European Economic Research (ZEW), Mannheim; CHI Research Inc.; Harvard University; and Centre for European Economic Research (ZEW), Mannheim respectively. Research grants from the Sloan Foundation and the ZEW are gratefully acknowledged. The authors thank Ulrich Schmoch of the Fraunhofer Institute for Systems Technology and Innovation Research, Karlsruhe, for help in accessing German citation counts. Please send correspondence to Frederic M. Scherer, John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138

I. Introduction

In an analysis of inventions emerging during the development of computed tomography (CAT) scanners, Manuel Trajtenberg (1990) found that patents covering inventions of greater social value were cited significantly more frequently in subsequent patents than inventions of lesser value. We confirm Trajtenberg's results for a much wider sample of U.S. and German patented inventions on which profitability (i.e., private value) information was obtained.

II. The Surveys

The sampling approach underlying our analysis was based upon two phenomena identified in earlier work with patents. See e.g. Lanjouw, Pakes, and Putnam (1996). First, many nations require patent holders to pay periodic renewal or maintenance fees to keep their patents in force. The longer fees are paid, the higher is a patent's implied value. Second, inventions originating in one national jurisdiction are in a sizeable fraction of cases also patented in other nations. The greater the invention's presumed value, the more likely patenting abroad is.

To exploit these two properties, we focused on German patents resulting from applications with a German priority date of 1977, and hence with a full-term expiration date falling within 1995.¹ Germany was chosen because it is one of the leading industrialized nations, because its patent examination system is particularly rigorous, and because its patent maintenance fee system is one of the most progressive in the world. Fees keeping a patent in force rose from DM 50 for the third year to DM 2,700 for the 18th year, cumulating to DM 16,075 over the full 18 years of life. From 57,720 applications with a 1977 priority date, 23,834 patents were actually issued. Of these, 4,349, or 18.2 percent, were renewed to full term expiration in 1995.

The sample analyzed here began with 1,431 full-term patents of German origin and 896 patents originating in the United States. Because not all holders of German-origin patents could be located, 1,352 recipients of German-origin patents were successfully contacted. From a stratified initial sample comprising 80 percent of the U.S.-origin German patents, links were successfully made to U.S. priority applications and from there to 485 parallel U.S. patents, which were the basis of a U.S. survey.

¹ The maximum patent life was increased from 18 to 20 years beginning with 1978 priority applications

Each sampled patent owner was contacted during 1996 (the year following patent expiration) and asked to answer a single question, which in the U.S. survey version was phrased as follows:

If in 1980 you knew what you now know about the profit history of the invention abstracted here, what is the *smallest* amount for which you would have been willing to sell this patent to an independent third party, assuming that you had a bona fide offer to purchase and that the buyer would subsequently exercise its full patent rights?

This asset value approach (consistent with the view of patents as intellectual property) was chosen following pilot interviews over a question asking for the accumulated flow of profits. Respondents in the German survey were asked to locate their patent's value within five categories ranging from "less than DM 100,000" to "more than DM 5 million." U.S. survey respondents were given a broader range of eight categories, from "less than \$100,000" to "more than \$100 million." Altogether, there were 772 useable responses covering the inventions of German origin and 222 useable responses in the survey of U.S. patents. Among the German patents, 12.9 percent were placed in the highest value category. Among the linked U.S. patents (thrice winnowed -- once for yielding a U.S. patent, again for yielding a German patent, and a third time for being renewed to full term in Germany), 22, or 9.8 percent, were placed in the "above \$100 million" category. Statistical distribution properties of the survey responses are analyzed in Harhoff et al. (1997).

III. U.S. Patent Citation Link Results

An exhaustive search of citations to sample patents contained in other U.S. patents subsequently issued (i.e., between 1978 and 1996) was made using the extensive data base of CHI Research, Inc. See e.g. Narin (1995). Twenty of the original U.S. survey responses lacked U.S. patent numbers required for a citation search; ten were received too late to be included in the search; and three were eliminated after linking because explanations accompanying the survey responses revealed that the respondent did not understand the question. Thus, the analysis here focuses on 192 useable citation-data-linked patents.

The number of citations recorded ranged from zero (nine cases) to 169, with a mean of 15.91, median of 9.0, and standard deviation of 20.83. The average number of citations to all U.S. patents issued in the same years as the sample patents was 6.02, or, when the universe figure is adjusted for average citation frequencies in U.S. Patent Office classes identical to those of the sample patents, 6.94. The differences between the sample and universe means are significant at better than the 0.01 level. Thus, U.S. patents selected on the basis of full-term German counterpart renewal are substantially more frequently cited than run-of-the-mill patents.

Figure 1 displays average citation frequencies for the eight economic value groups into which respondents classified their patents, with the number of observations per group recorded in parentheses. In every value group, the average citation frequency for full-term German counterpart patents exceeds the universe average of 6.02. Within our full-term sample, two broad groupings appear. The most profitable patents -- those valued at \$20 million or more -- are visibly more highly cited than less valuable patents, i.e., 29.6 times on average for the former, compared to 13.1 for the latter. An F-ratio test reveals the two groups to be significantly different; $F(1,190) = 18.66$. A less-structured test for differences among the eight classes is also highly significant; $F(7,184) = 3.12$.

Although the group citation averages in Figure 1 do not rise strictly monotonically with estimated patent value, nonlinear regression analysis supports the existence of a generally positive relationship. Six of the eight value groups were assigned values equal to the geometric mean of their bounds; the mean value of the most valuable patents was estimated (on the basis of interviews) to be \$250 million and the mean of the least valuable patents to be \$50,000. The regressions, with and without dummy variables taking into account the eight main classes of the International Patent Classification (IPC), are presented in Table 1. (Standard errors are enclosed in parentheses.) In OLS regressions, zero values of the citation count dependent variable are assumed to be 0.3. Negative binomial regressions avoid this arbitrary zero-value convention and also permit closer comparison with later German patent results. In all cases, the economic value coefficients (interpretable as elasticities) are statistically significant, with t-ratios of 3.6 or higher. The relationships exhibit strongly diminishing returns. In regression 1a, for example, the average economic value associated with an average citation reaches \$1 million at roughly 10.5 citations, i.e., slightly above the median number of citations for all sample patents. At higher citation levels, the corresponding economic value is greater than \$1 million, the more so, the more valuable the patent. Among the seven dummy variables, only the coefficient for IPC class G (corresponding approximately to the scientific and professional instruments sector under the U.S. Standard Industrial Classification) was statistically significant (and positive).²

The relatively low R^2 values for all regressions reveal that the citation - value relationship is quite noisy. Aggregation of citations and economic value estimates from the individual patent level, as in Table 1, to the company level was possible for 48 companies with complete responses. The analogue of regression 1a in Table 1 is:

² In the U.S. citation universe, patents with that IPC code also had unusually high citation frequencies. Regressions adjusting individual patent citation counts for universe average citation frequencies had citation - value relationships similar to those in Table 1.

$$(1) \ln \text{CITES} = 2.336 + 0.317 \ln \text{VALUE}; \quad r^2 = 0.285; n = 48;$$

$$(.270) \quad (.074)$$

with standard errors in subscripted parentheses. The higher r^2 here shows the effect of pooling 3.11 patents on average per company, with a range of from 1 to 10 patents.

Company-financed research and development outlays in 1976 were disclosed by 46 companies on which complete U.S. citation counts were also available. With citation counts aggregated to the company level, the OLS regression is:

$$(2) \ln \text{CITES} = 1.093 + 0.580 \ln \text{R\&D}; \quad r^2 = 0.240; n = 46.$$

$$(.645) \quad (.156)$$

A positive and significant relationship emerges between recorded citations and the R&D spending that gave rise to the cited patents (as well as to many more patents without full-term German counterparts).

The data at hand made no distinction between citations to a firm's own patents and other companies' patents. It is possible that companies with more patents engage in more self-citation. For a crude test of this hypothesis, a count was made of the total number of U.S. patents issued in 1978 to sample companies on which research and development expenditure data were also obtained. Total recipient company patent counts, varying from seven to 503, were available for 157 of the 192 patents in the citation count sample. The relationship between the logarithms of total patent counts and citations was negative, contrary to hypothesis, but statistically insignificant. When the total patent count variable (in logarithms) was added to the analogue of regression 1a in Table 1 (with $N = 157$), the regression coefficient for $\ln \text{VALUE}$ changed from 0.113 to 0.116. The total patent count coefficient was negative and insignificant ($t = 1.55$). Thus, there is no support for the hypothesis that companies receiving large numbers of patents had more highly cited patents.

IV. German Patent Citation Link Results

Using the data banks of the Fraunhofer Institute for Systems Technology and Innovation Research in Karlsruhe, Germany, a similar link was effected between patents of German origin on which economic value estimates were obtained and the German patent citations to those patents. The economic value survey for German patents proceeded in two stages. In the first stage, telephonic and facsimile communications elicited *Gedankenexperiment* value estimates in five discrete value categories. In the second stage, interviews were conducted with the owners of most of the patents whose estimated value exceeded DM 5 million, yielding inter alia more precise point estimates of those inventions' values. See Harhoff et al. (1997).

Successful citation links were achieved for all 772 of the patents on which German survey responses were secured. A tally of citation counts was also made for all 11,364 German patents of German origin resulting from applications with 1977 priority dates. This research yielded the initially surprising finding that German patents cite prior German patents much less frequently than U.S. patents cite prior U.S. patents. For the universe of German-origin patents with 1977 priority dates, the average number of subsequent citations was 0.528 per patent (compared to six for corresponding U.S. patents). The maximum number of recorded citations was 15; 68.41 percent of the German universe patents received no citations at all.³ There are two or possibly three reasons for this difference. First, roughly four times as many patents were issued per year during the 1980s and early 1990s in the United States as in Germany; therefore, there were more U.S. patents to cite any given U.S. predecessor. Second, citation practice in Germany is probably more conservative than in the United States. Third, beginning in 1978, many of the most important German-origin inventions were covered by Europatents rather than by domestic German patents. Relatively important patents may make more citations, especially to important predecessors, than less important patents. This hypothesis will be tested with Europatent data not yet available.

Patents renewed to full term and on which economic value estimates were obtained through our German survey were more likely to be cited and had more citations on average than patents allowed to lapse before their maximum 18-year term. The average number of citations per survey response patent was 0.703, 40 percent more than the non-full-term German cohort average of 0.498. The difference in means is highly significant, with $t = 5.48$. Among the survey patents, 60.36 percent had zero citations, compared to 69.37 percent for the non-full-term cohort. In a binomial proportions test, the difference is highly significant, with $t = 4.95$.

Figure 2 displays the average citation frequencies (including zero-citation cases) for survey response patents grouped by their estimated economic value. For patents originally valued at more than DM 5 million, finer disaggregations from subsequent interviews are used. Also shown is the average for non-full-term cohort patents. The two most valuable inventions are grouped together for reasons of confidentiality; disaggregation would not change the implications. As with the U.S. patents, three regimes emerge from the array. Early-expiration universe patents are less cited than all but the least valuable full-term patents. The second- and fourth-highest average

³ Four of the eight universe patents with ten or more citations had been renewed to full term, and hence were included in our survey frame, but their owners chose not to participate in the survey. From inspection of those patents, there is reason to believe that very high economic values would have been reported. Among the total of 659 full-term patents on which no response was obtained, the average number of citations was 0.765; 63.3 percent had zero citations. The mean number of citations conditional upon non-zero citation was 2.083.

citation frequencies appear among high-value inventions in the DM 10-20 and 40-80 million categories. By far the highest citation frequencies are for the two patents with the greatest reported economic value.

This visual implication is confirmed by regression analysis of individual patent citation counts on reported economic values. Due to the high incidence of zero values and the low citation counts, ordinary least squares regression is inappropriate. Poisson regression is also ruled out, since the variance of the citation counts is roughly twice the mean count. Therefore, a more flexible negative binomial formulation is used, with economic VALUE measured (in millions of DM) as the geometric mean of reporting class interval bounds for first-stage survey responses (with a minimum value of DM 30,000 for the "less than DM 100,000 group" ⁴) and as the mean point estimate for second-stage interview responses. The resulting regressions, with the number of citations as dependent variable and the economic value estimates (along with seven IPC dummy variables) as explanatory variables, are presented in Table 2.

Several implications stand out. First, in every regression the regression coefficients on \ln VALUE (interpretable as elasticities) are significantly positive, with t-ratios in the full-sample regressions of 2.72 and 3.09. Second, the full-sample elasticities lie within a standard error of those obtained with the completely different U.S. sample of Table 1, despite the much smaller average number of German citations. Third, when the sample is limited to 66 patents with value estimates exceeding DM 5 million, higher elasticities with substantially larger standard errors emerge.⁵ Fourth, the seven International Patent Classification dummy variables together are statistically significant in a Chi-square test. As in the U.S. data, the "G" class had particularly high citations, as did (unlike the U.S. results) the "D" class (including textiles and paper goods). But finally, the pseudo- R^2 values are even more modest than in the U.S. regressions, revealing again that the value - citation relationship is quite noisy.⁶

⁴ Presumably, the least valuable full-term patents were expected to recover their renewal fees of DM 16,075 plus application and legal fees of roughly the same magnitude.

⁵ This was also true for the U.S. sample. When binomial regressions were confined to 97 patents valued at \$1 million or higher, the VALUE coefficient was 0.207 (0.059) without IPC dummy variables and 0.182 (0.063) with dummy variables.

For regression 2b with 66 high-value patents, one could not confidently reject the hypothesis that the citations were Poisson-distributed. The \ln VALUE coefficient in a Poisson regression with IPC dummies was 0.387 (.100).

⁶ Pseudo- R^2 values were computed by squaring correlations between the actual and regression-predicted numbers of citations (in absolute levels, not logarithms).

V. Conclusion

In sum, our results from two wide-ranging surveys, one in Germany and one in the United States, support Trajtenberg's conclusion that patents with greater economic value are more heavily cited in subsequent patents. Our evidence suggests at least a two-stage escalation of economic values and citation counts. First, patents that are renewed to full term expiration in environments such as Germany with highly progressive annual maintenance fees are more valuable and more highly cited than run-of-the-mill patents allowed to expire before running to full term. Second, within the relatively exclusive cohort of full-term patents, citation frequency rises with economic value, although with considerable noise in the relationship. The most cited patents turn out to be very valuable indeed, with a single U.S. citation implying on average more than \$1 million of economic value. These findings provide new support for research seeking to overcome the limitations of simple patent counts as a measure of innovative contribution by acquiring data on citation frequencies and the number of nations in which patent protection is sought. We have also developed a methodology, useful primarily after roughly two decades and hence for historical analyses, to identify the most valuable inventions in any much larger cohort of inventions on which patents have been issued. Through extensions of that methodology, it should be possible to achieve more detailed insights into the magnitude and distribution of economic returns to investments in technological innovation.

References

- Harhoff, Dietmar, Scherer, F. M., and Vopel, Katrin, "Exploring the Tail of Patent Value Distributions"; ZEW Working paper no. 97-30, forthcoming November 1997.
- Lanjouw, Jean O., Pakes, Ariel, and Putnam, Jonathan, "How To Count Patents and Value Intellectual Property"; NBER Working paper no. 5741, 1997.
- Narin, Francis, "Patents as Indicators for the Evaluation of Industrial Research Output"; *Scientometrics*, vol. 34, 1995, pp. 489-496.
- Trajtenberg, Manuel, "A Penny for Your Quotes: Patent Citations and the Value of Innovations"; *RAND Journal of Economics*, vol. 21, 1990, pp. 172-187.

Table 1: Regressions of U.S. Patent Citations on Value (N = 192)

Variable	OLS		Negative Binomial	
	1a	2a	1b	2b
Constant	2.076 (0.089)	1.845 (0.273)	2.682 (0.074)	2.267 (0.228)
ln Value	0.119 (0.033)	0.137 (0.034)	0.105 (0.027)	0.136 (0.028)
IPC Dummy Variables	-	diverse	-	diverse
R ² or pseudo R ²	0.063	0.120	0.064	0.100

Table 2: Negative Binomial Regressions of German Patents Citations on Value

Variable	All Full-Term Patents		Full-Term Patents Valued Above DM 5 mil.	
	1a	2a	1b	2b
Constant	-1.534 (0.435)	-0.703 (0.475)	-4.606 (2.329)	-6.618 (2.170)
ln Value	0.090 (0.033)	0.105 (0.034)	0.272 (0.139)	0.345 (0.129)
IPC Dummy Variables Chi-squared (df)	-	diverse 18.13 (7)	-	diverse 10.87 (7)
N	752	752	66	66
Chi-Squared (df)	743 (1)	24.45 (8)	3.82 (1)	11.80 (1)
log L	-859.2	-850.6	-86.7	-82.8
pseudo-R-squared	0.012	0.030	0.176	0.309

Figure 1
 U.S. Patent Citation Frequencies by Value Group

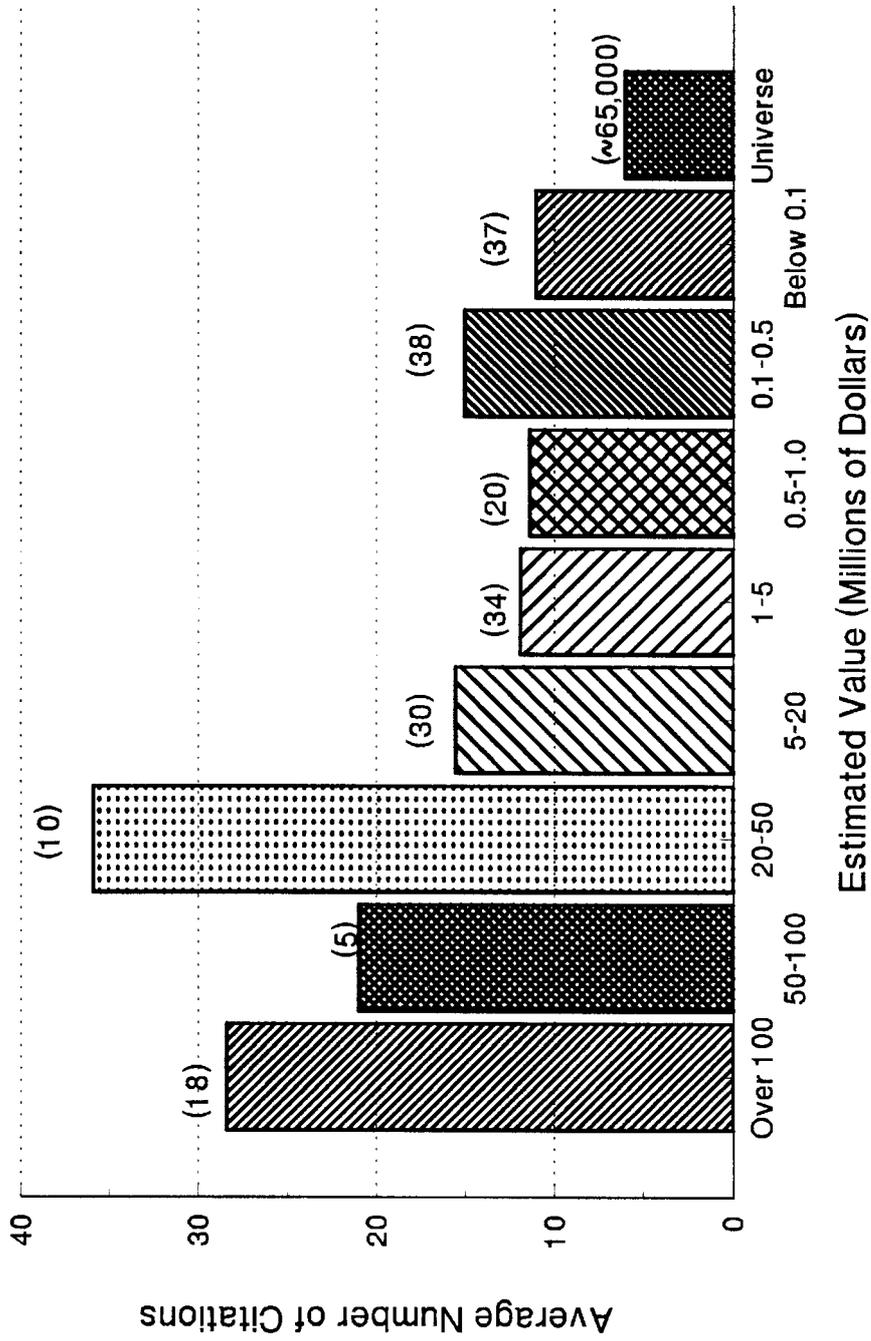


Figure 2
 German Patent Citation Frequencies by Value Group

