

Patent Citations: A Review

by

Xiaoyan Cai

Student No.: 2442592

**Major paper submitted to the
Department of Economics of the University of Ottawa
in partial fulfillment of the requirements of the M.A. Degree**

Supervisor: Professor Gamal Atallah

Ottawa, Ontario

January 2003

ABSTRACT

This paper reviews the main uses of patent citations in the literature on innovation, such as knowledge spillovers, firm market value, assessment of the value and quality of innovations, and also the use of citations in some other fields. Citations are a valid but noisy indicator of knowledge spillovers. Moreover, citations provide a better explanation of the market value of the firm than simple patent counts. Finally, forward citations are the most relevant indicator of the value of innovation. Some methodological issues, such as the citations made by the examiner, self-citations, and the truncation problem are also discussed.

TABLE OF CONTENTS

Chapter 1 Introduction	1
Chapter 2 Fundamentals of patents and patent citations	3
2.1 Patents' use in research	3
2.2 Patent citations	5
Chapter 3 Patent citations as an indicator of knowledge spillovers	7
3.1 Control, Generality and Appropriability	7
3.2 Evidence of localization	10
3.3 Non-localization in LCD technology	15
3.4 "Generality" and "Importance"	17
3.5 Summary	20
Chapter 4 Using patent citations in market value assessment	22
4.1 Market value functions	22
4.2 A non-parametric series estimator to deal with data shortages	27
4.3 Using expected weighted patent counts to eliminate the truncation bias	29
4.4 Summary	30
Chapter 5 Using patent citations in innovation value assessment	32
5.1 Citations and the value of innovation	32
5.2 Citations and innovation quality in a latent model	35
5.3 Summary	38
Chapter 6 Patent citations in other fields	39

Chapter 7 Avenues for future research	41
7.1 Patent citations made by the examiner and self-citations	41
7.2 Truncation of the patent citation data	42
7.3 The value of the cited patent	43
Chapter 8 Conclusions	45
Appendix	46
Bibliography	60

Chapter 1

Introduction

In the past few decades, economists have become increasingly interested in studying how R&D activities influence the market value of firms, how to evaluate the value of innovations and how to determine the knowledge spillovers between different companies or different geographic areas. Although different approaches have been adopted to tackle these questions, patents and patent citations are the most used indicators.

A patent is defined as an exclusive right to the inventor for the commercial use of an innovation within a limited period of time. Patent citations are the references to the previous inventions in the same field or in different fields of technology that the patent inventor cited. Most studies find a strong relationship between patent citations and innovation value, knowledge spillovers and firm market value.

In this paper, we will focus on how economists use patent citations data to analyze knowledge spillovers, market value and the value of innovation. The paper is divided into eight chapters. In Chapter 2, some basic terminology regarding patents and patent citations is presented. Chapter 3 examines how patent citations are used in studying knowledge spillovers. Chapter 4 will analyze market value. Chapter 5 will focus on measuring the value of innovations using patent citations. These three chapters go through some representative studies in detail, making comparisons in terms of the approaches adopted, the data used, and the results derived. Chapter 6 analyzes some other areas in which patent citations have been used. Chapter 7 reviews the

weaknesses in the current studies of patent citations and examines future developments.

Chapter 8 concludes. The appendix includes all the tables and figures mentioned in this paper.

Chapter 2

Fundamentals of patents and patent citations

2.1 Patents' use in research

As Tranjtenberg put it (Tranjtenberg, 1990:183): “patents are the only observable manifestation of inventive activity which have a well-grounded claim for universality.” A patent is protected from commercial exploitation by others without permission. In order to get the patent, the innovation must be non-trivial, new and different from other contemporary technology; moreover, it must have a commercial potential value (Jaffe et al., 1993).

A patent is defined as an exclusive right to the inventor for the commercial use of an innovation within a period of time, generally 20 years. Except laws of nature and inventions offensive to public morality, all inventive activity can be patented. In exchange for disclosing the innovation in the patent published by the Patent Office, a monopoly is granted. The monopoly begins on the date of issue, and expires at the end of the period. On average, two years separate filing from issuing, although this delay may vary considerably. A maintenance fee every four years is necessary to keep the patent in force. It is possible for the applicant to patent an innovation in several countries.

The process of obtaining a patent varies among countries. Generally the process can be divided into six steps:

1. The filed application contains a detailed description of the innovation and of claims concerning the scope of protection requested by the applicant;

2. The Patent Office starts to search the previous filed innovations that affect the current application;
3. The application is published;
4. The applicant requests an examination of the claims;
5. The Patent Office carries out the examination and raises objections;
6. The applicant amends the application and the patent is granted.

In economic research, patents are found to be most useful in evaluating the output of R&D (Hall, 1998). The basic assumption here is that the more patents a firm has, the more effective the R&D resources the firm has invested are.

Schmookler (1966), Scherer (1965) and Griliches (1984) have used patent data for studying economic growth, firms' size and productivity. Griliches (1984) is one of the first to point to patents and patent citations as indicators of innovation. Scherer (1965) argued that the use of simple patents counts is problematic: most patents contain important information about the innovation disclosed in them, but not all patents carry significant value. Therefore, a more accurate measure is needed in the study of measuring patent values.

Patent citations have not been widely used in economic research until the late 1980s, mainly due to a lack of availability of patent citation data. Starting from the late 1980s, patent citation data became more available with the aid of computerization.

2.2 Patent citations

Patent citations are the references to the previous knowledge or technology in the same field or in different fields of technology. When a patent is granted, the list of these references is detailed in the documents together with other details about the inventor, the invention and the organization which receives the patent's property right.

Patent citations can be made either by the patent applicant and his/her attorney or by the patent examiner (Hall et al, 2000). Patent citations play an important role because inventors use patent citations to state the status of previous technology upon which the current invention builds and the patent examiner uses patent citations to identify previous innovations which may be missed or concealed in the application, or to oppose the application altogether (Jaffe et al., 1993).

The number of citations received by a patent varies dramatically from one patent to another. Some patents have been cited for thousands of times while some other patents received no citation at all. The distribution of patent citations is highly skewed (Hall et al., 2000), indicating that only a few patents are very valuable, while most patents have a very low value. It has been widely consented among economists that citations data does convey the importance of the innovation disclosed in a patent (e.g. Jaffe et al., 2000). Therefore, it is reasonable to think that the more citations a patent gets, the more valuable the innovation embedded in that patent is.

In recent decades, a number of areas have been studied using citations as an indicator of knowledge spillovers. Important contributions have made by Jaffe et al. (1993), who use citations as an indicator of spillovers. In the study of economic impact using citation data, Trajtenberg (1989, 1990) found that there is a strong correlation

between patent weighted counts based on citations and the value of innovation, which proves that citations can be used as a measure of a patent's quality.

One weakness of citations data is that they contain a lot of "noise", which can be defined as the citations added by the patent examiner or inventor's attorney that are not known by the inventor. Jaffe et al. (2000) found that a large number of citations have been added by the patent examiner when the examiner carries out the examination of the patent application. Since the inventor is unaware of them, these citations cannot be a good indicator of spillovers. Jaffe et al. (2000) showed that a large fraction of citations (about one half) does not correspond to any apparent spillovers. These results are consistent with the notion that citations are a noisy signal of spillovers. At the same time, however, the results also show that about one quarter of citations correspond to a fairly strong spillover, or some degree of spillovers.

Citations delimit the property right granted by the patent and identify the previous technology which is not covered by the citing patents, in addition to measuring the importance of the technology contained in the cited patents. Researchers have used citations as an important indicator to measure and evaluate knowledge spillovers (e.g. Jaffe et al., 1993 and Stolpe, 1998), market value (Shane, 1997, Hall et al., 2000 and Bloom, 2001) and the value of innovation (Trajtenberg, 1990 and Lanjouw and Schankerman, 1999).

Chapter 3

Patent citations as an indicator of knowledge spillovers

Economists have found that spillovers constitute an important input in generating new ideas (Caballero and Jaffe, 1993 and Schmookler, 1966). A number of issues arise in the study of knowledge spillovers as for example: Is proximity an advantage for nearby firms or for firms in the same country? or Can we track spillovers? Finding answers to these questions is crucial in many decision making processes.

A large literature has focused on these questions. Krugman (1991) emphasizes that knowledge flows are invisible and that they do not leave any paper trails. In contrast, Jaffe and Trajtenberg (1993), Caballero and Jaffe (1993) and Verspagen (1997) believe that knowledge does leave a paper trail in the form of patent citations, since patents contain detailed geographic information on inventors. By studying the relationship between citations and spillovers, we will know where the trails lead and be able to test to what extent spillovers are localized.

3.1 Control, Generality and Appropriability

The most popular method in the study of localization is to construct “control” samples besides the citing patents. “Control” is defined as “the patent whose grant date is as close as possible to the grant date of citing patent and the distribution across time and technological areas is essentially identical to that of the citation data set” (Jaffe et al., 1993: 582). In these samples, “Controls” are not citations but have the same

temporal lag and technology distribution as the citing patents. "Controls" are constructed to help us make a comparison of geographic matching probabilities between the citing/non-citing patents and originating patents. We can then test whether citing patents are more localized than the non-citing ones. The expected result is that the citations matching frequency is significantly greater than the control frequency, which would imply that spillovers are localized. There are at least three good reasons for using "control".

First, a "control" sample provides a non-citation conditioned probability which gives a reference value against which to compare the frequency with which citations match. Second, using control can help avoid some mistakes. For instance, if one mistakenly counts a foreign citation into local matching between citations and originating patents, this will decrease the probability of the local matching. However, because the control method is adopted, and the wrong operation also appears in the control matching group, the probability of matching between the control and the originating patents is also decreased, so the result will not be biased. Thirdly, in terms of the pre-existing pattern of geographic concentration of technology, consider the example mentioned in Jaffe et al. (1993). A large fraction of citations to Stanford patents coming from Silicon Valley will be attributed to localization of spillovers. Because Stanford happens to be located in Silicon Valley, we might attribute the citations to localization. If Stanford was not located in Silicon Valley, there would still be a large number of citations to Stanford patents coming from Silicon Valley. In this case, the probability of matching between control and the originating patents and that between the citations and the originating patents are both higher than the real

probability, but since the probability is compared relatively, the resulting error does not bias the results.

Most of the studies also study the factors affecting localization by running probit estimations. The two most commonly used explanatory variables are “generality” (e.g. Hall et al., 2001; Jaffe et al., 1993) and “appropriability” (Jaffe et al., 1993; Stolpe, 1998). “Generality” is defined as one minus the Herfindahl index; it is used to test the extent to which the citing patents are diverse in terms of technology. “Generality” can be measured using the following equation (Jaffe et al., 1993):

$$G_i = 1 - \sum_k \left(\frac{C_{ik}}{C_i} \right)^2 \quad (3-1)$$

where G_i is the generality for patent i , C_i is the total citations received by patent i , C_{ik} is the number of citations patent i receives from subsequent patents whose primary patent class is k . When k equals one, G_i equals zero when all citations are received from patents in the same class, and when G_i approaches unity, citations are very dispersed. Appropriability relates to self-citations; “a high proportion of self-citations is the evidence of relatively successful efforts by the original inventor to appropriate the invention” (Jaffe et al., 1993).

3.2 Evidence of localization

Jaffe et al. (1993) identified, for each citing patent, all patents in the same patent class¹ with the same application year, and then chose a control patent from this set whose grant date was close to that of the citing patent. Thus, they derived a set of control patents for each citing patent whose distribution across time and technological areas is identical to that of the citing patents set. The idea of using this method is to compare two probabilities in two situations: the first is the probability of geographic match for two patents, conditional on a citation link and the technological nature and timing of the citation. The other probability is the probability of “control” patents matching geographically the cited patents, conditional only on the technological nature and timing of the citation. If the citation match frequency is significantly higher than the control match frequency, then citations are localized. If knowledge spillovers are localized, the authors believe that “citations should come disproportionately from the same State or metropolitan area as the originating patent” (1993:579). One can explain this idea in terms of probability: if there is no localization, the probability of citing patents in the same area should be the same as that of citing patents in other areas; conversely, if there is localization, the probability of citing patents in the same area should be larger.

Jaffe et al. (1993) ran a probit estimation with geographic match/no match between the originating and citing patents as the dependent variable on independent variables which include:

¹ The IPC (International Patent Classification) divides all technological fields into 8 fields, 120 classes, 628 subclasses and approximately 69000 groups.

- the *citation lag*, which is defined as the difference between the application year of the citing patent and the application year of the cited patent;
- dummy variables for *top corporate* and *other corporate*² originating patents;
- interactions of the lag with these dummies;
- a dummy variable equal to unity if the citation has the same class as the originating patent;
- a dummy variable equal to unity if the control patent corresponding to this citation matches the originating patent geographically;
- two variables relating to generality and appropriability of inventions which are the generality of the originating patent and the fraction of the originating patent's citations that are self-citations.

They began the study with two cohorts of “originating patents” which consist of 1975 patent applications and 1980 applications, respectively. In each cohort, they included the patents granted to U.S. universities and two samples of U.S. corporate patents to match the university patents by grant date and technological distribution. The 1975 cohort contains about 950 patents that had received a total of about 4750 citations by the end of 1989; the 1980 cohort contains about 1450 patents that had received 5200 citations.

In their study, the raw dataset has a higher probability of receiving citations for both 1975 and 1980 university patents. And according to table 3-1, the average citation

² The “top corporate” consists of patents granted to the 200 top-R&D-performing firms in the U.S. The “other corporate” contains patents assigned to U.S. corporations that are neither universities nor “top corporate” (Jaffe et al., 1993).

lag for 1980 is obviously shorter than that of 1975 patents: the average lag between the originating application year and the application year of the citing patent is 6.5 to 8 years for the 1975 cohort, and is over 4 years for 1980 cohort. Universities have the lowest rate of self-citation in both the 1975 and 1980 cohorts (5.6 and 8.9, respectively), while top corporate has the highest rate of self-citation in both cohorts (18.6 and 24.6, respectively).

Figure 3-1 contains two graphs, which provide some additional information on the distribution of lags of citations for 1975 and 1980 between originating and citing patents. Looking at the 1975 and 1980 cohorts, we can see that citations for both the 1975 and 1980 cohorts are few in the early years, reach a maximum after 3 years, and then tail off. The difference between the two graphs is that the citations rate of University patents is particularly higher for 1975 in the early years than that for 1980. From figure 3-1, it is reasonable to believe that the drop in citations at the end is partly due to the cut off of the observations in 1989.

In order to get some general idea about localization, Jaffe et al. (1993) also compared the location match between citations and controls to the original patents in major SMSA³ areas. The fraction of citations to 1975 University patents Madison/Wisconsin is higher than that of controls, suggesting a presence of localization. Also, we observe a high percentage of citations to patents from SMSA for the location New York/NJ/CT⁴.

Table 3-2 presents how citations match the originating patents geographically before and after excluding self-citations. We first look at 1975 at the country level.

³ Metropolitan Statistical Areas.

⁴ NJ stands for New Jersey, CT stands for Connecticut

Before excluding self-citations, the percentage of overall citation matching is 6-7% higher than the controls. After excluding the self-citations, the citation matching frequencies are all decreased by 2%, and more significant, they are lower than that of the control for top corporate. Patents at the State level, rate of citations coming from the same State with original patents is about 10% for University, 19% for top corporate and 15% for other corporate. After excluding self-citations, citation matching frequency was cut to 6% for University, 6.8% for top corporate (which is the same as the matching frequency of control), and over 10% for other corporate, suggesting a higher matching frequency than control. At the SMSA level, 9% to 17% of total citations are localized. These results drop significantly after self-citations are excluded, but the difference is still significant.

The results for the citations of 1980 patents are even more significant than for the 1975 patents. At each geographic level, the citations are significantly more localized than the controls. For example, at the country level, after excluding self-citations, the overall citations matching is higher than control by 9% to 11%. This number is much higher than that for the 1975 cohort. And also, we can observe a decline of about 3% to 6% for controls matching at the country level, suggesting an increasing share of patents are assigned to foreigners.⁵

We can summarize the results as follows: there is a clearer pattern of localization for 1980 patents at the country, State and SMSA levels compared to the 1975 patents. At the State and SMSA levels, University citations are as localized as the firms' citations. Citations of 1975 patents have the same pattern but localization is

⁵ This conclusion is drawn because matching frequency for controls is not conditioned on citation.

weaker. There is no localization at either the country or the State level for top corporate, but there is significant localization for SMSA.

In terms of exploring the factors affecting the probability of localization, the results of the probit estimation for 1975 citations in Jaffe et al. (1993)'s study imply that the localization of the corporate and University citations fades over time with the same speed. At the State and SMSA levels, there is also evidence that a lot of self-citations tend to be domestic. Conversely, the regression results of citations of the 1980 patents are not very satisfying: the sign of the time lag coefficient changes, partly because the citation span is too short to show the lag effect. After rerunning the regression with 1975 citations excluding the citations granted after 1984, Jaffe et al. (1993) obtained the same results as for 1980. Similarly, we could presume that the results will be also the same as the results of 1975 if the citations of 1980 patents were added, since we previously mentioned that the citation lag is shorter for the 1980 cohort.

Based on these results, Jaffe et al. (1993) concluded that there is localization of knowledge spillovers in the form of citations. Moreover, the 1980 citations are obviously more localized than the 1975 citations⁶. They also found that geographic localization fades over time. Besides self-citations for corporate which are more localized, they did not find significant differences between citations of University and corporate patents in their study. Neither did they find any evidence showing that the probability of citations coming from a given geographic location conditional on patent class is different from the unconditional one.

⁶ This was mentioned when we explained table 3-2, the citations matching is higher than control by 9%-11%, which is much higher than that for the 1975 cohort, suggesting a more significant localization.

3.3 Non-localization in LCD⁷ technology

In Stolpe's (1998) study, controls were also used. Stolpe used data from U.S. patent class 349⁸ covering 1398 patents filed between 1975 and 1995. The patents involve 2116 different inventors and are owned by 236 different assignees.

Stolpe analyses two discrete choice models, one for the determinants of citations choice and the other for the determinants of localization among patent citations. First, he used citations to the 178 original patents filed during 1980-1983 which were chosen from subsequent patents in class 349 filed up to 1995. Second, for each of the 645 documented citations, the citing patent was paired randomly with a non-cited patent from the same group of 178 original patents. He thus generated a data set of citing and non-citing patent pairs with equal numbers and identical temporal lag distributions. In the first model, he included some dummy variables to look for the determinants of citation choice, and also tested whether the dummy means for the citations equal those for the non-citations. Since self-citations are assumed not to constitute a spillover, they have been removed from the data set, so only 618 citations and 629 non-citations were used. In the regression, Stolpe used *local matches* as the dependent variable, dummies for *control sample match*, *subclass equality*, *examiner equality*, *assignee equality*, *prior cooperation*, and some lags of years and measures of *appropriability* and *basicness* (or *generality*) as explanatory variables. Since in the regression of determinants of citation choice, the full set of data has not been included, a likelihood function needs to take into account the probability of inclusion in the sample.

⁷ Liquid Crystal Display.

⁸ Liquid Crystal Cells Element and Systems.

In contrast to the other studies, Stolpe finds that the diffusion of knowledge among inventors in Liquid Crystal Display technology does not appear to be localized. First of all, Stolpe showed through descriptive statistics on patent flows that the Gini coefficient for countries increases with the number of countries, suggesting that inventive activity of LCD is concentrated in a small group of countries. This trend can be explained by an increase in patents granted from the late 1980s to the early 1990s. In the year 1993 alone, there were 102 patents granted in Japan, compared to only 36.5 patents granted in the U.S.. South Korea has obtained 15 patents in 1993, while Germany and Switzerland have fallen behind. This suggests that Japan has dominated patenting since the beginning of the period. Before regressing the citations choice model, Stolpe compared the means of dummy variables used in the model between citations and non-citations, and found the means of neighboring assignees, distant assignees in one country and assignees in different countries to be significantly different between the citations and non-citations.

Table 3-3 shows the regression results of the citation choice model. The results suggest that only *basicness* and *subclass equality* have a strong impact on citations choice; *appropriability*, *prior cooperation* and dummies for geographical distance between inventors do not have any significant influence on citations choice. Moreover, knowledge diffuses less in the U.S. than in Japan.

In order to know the determinants of spillover localization, Stolpe regresses local matches among inventors on the same set of explanatory variables except for the distance dummies. The dummies for *control sample match*, *appropriability*, *subclass*

*equality, assignee equality and prior cooperation*⁹ have a very significant and positive impact on geographic localization, while *basicness* and *examiner equality* (which is defined as both patents having been cleared by the same examiner) *have* a negative impact.

Based on Stolpe's study, there is no strong evidence of localization of spillovers in the field of LCD technology, since the probability of citation does not show a strong correlation with the distance dummies. Moreover, in both the citation choice model and the localization model, subclass equality is important, suggesting that regions have developed their own distinct pattern of technology specialization in the LCD field. In determining geographic localization, prior cooperation is a significant factor, which suggests that collaboration between the inventors of later patents and the inventors of earlier patents does play a role in localization.

3.4 "Generality" and "Importance"

Not all studies use "control" sample. Jaffe et al. (1997) mainly used the frequency and diversity of citations to measure the degree of "importance" and "generality" of the inventions and how these two variables changed over time. They investigated the problem of geographic localization of patent citations. To construct the degree of "importance", they derived citations made to each of the 55710 patents in the basic sample from patents granted between 1977 and 1994, and then call these citations count "importance". The definition of degree of "generality" is the same as above.

⁹ Prior cooperation indicates that any of the inventors of the later patent is linked to any of the inventors of the earlier patent by having been collaborators or colleagues at some third organization between the filing dates of the paired patents.

Jaffe et al. undertook simple analysis regressing “importance” and “generality” on 30 year dummies and 390 class dummies, and also the dummies for being the NASA or “other federal” sub-samples. These regressions were carried out for each of the five technical fields: Drugs/Medical, Chemicals, Electronics/Nuclear, Mechanical, and All Other. With respect to geographic localization, they focus on the patents whose primary inventor resides in the States where NASA¹⁰ patents originate, and refer to these states as “NASA complex”. They regress the fractions of citations which are from the same State, complex States, other U.S. States, and foreign countries respectively on dummies of cited year, patent classes, NASA (consisting of 3782 patents assigned to NASA between 1969 and 1993) and other federal (consisting of the 37939 patents assigned to the U.S. government between 1963 and 1993, excluding those identified as belonging to NASA) before and after excluding self-citations.

Before conducting the formal regression process, Jaffe et al. (1997) found evidence of geographic localization in their study: Virginia/Maryland/D.C., Alabama, Ohio and California have higher rates of NASA citation than the other States. Figure 3-2 presents the fraction of citations for different States.

Before adding a time effect, a simple regression based on class dummies shows that NASA patents are significantly more important in the field Mechanical, and less important in the Chemicals and Electronics/Nuclear categories. In terms of generality, NASA is more “general” in the category Mechanical.

After the year dummies are added, the coefficients of the *other federal* dummy turn out to be less “important” than those of *NASA* dummy, and NASA patents appear to be very “important” in the second half of the 1970s, but are not so different from

¹⁰ National Aeronautic and Space Administration.

other federal patents by the second half of the 1980s. In terms of generality, non-NASA patents are less “general” than the random sample¹¹ in the late 1960s, but are relatively more “important” from the late 1970s until the present.

Turning to geographic localization, the upper part of Table 3-4 shows that NASA patents and the other federal patents get fewer citations from the same States compared to the random sample, but get a higher rate of citations from “complex” States.¹² The corresponding regression on dummies for *cited year*, *patent classes*, and dummies for *NASA* and *other federal* and also a couple of States with NASA patents is shown in the lower part of Table 3-4. Again, NASA patents and the other federal receive most of the citations from “complex” states, but compared to NASA, the other federal seems to have a weaker pattern.

These results are obtained without excluding self-citations. Table 3-5 shows the results after self-citations are excluded. The same conclusion continues to hold, except that the fraction of citations from the same State was cut by a third, while the fraction from other “complex” States is almost unchanged. Similarly, NASA patents are less localized in the same State after the exclusion of self-citations, implying that self-citations are more localized, while the fraction from “complex” states is similar for all three groups, albeit with less localization for the other federal. Overall, there is evidence of localization of spillovers, and this conclusion holds even after the exclusion of self-citations.

¹¹ This sample consists of patents granted to corporations, to individual inventors, and also to the federal government. Federal patents are only about 3% of the universe of domestic patents. This sample contains 13997 patents for the period 1963-1993.

¹² “NASA complex” is a subset – the patents whose primary inventor resides in one of the States from which a significant number of NASA patents originate: California, Ohio, Texas, Alabama, Maryland, Virginia and District of Columbia.

3.5 Summary

In all of these studies, three factors have to be considered: technological categories, geographic location and time. First, some of the studies specify the category in order to study how the dependent variables behave among different fields. For example, in the case of Liquid Displayed Technology, the study focused on one technology, thus only one category was considered. Second, geographic location is an important factor in the study of the localization of spillovers. It can be specified at different levels, such as State, country and foreign country, or can be specified in certain areas, like “a complex of NASA”, for example. Third, time, measured as the lag between the citing patent and the cited patent, is also considered an important factor, since localization fades over time as the technology disseminates. In addition, basicness (generality) and appropriability are included in the regression model to consider how diverse the patent is across the technology fields and how localized the self-citation is.

The noise of citations as a measure of spillovers has been noticed in most studies. Trajtenberg and Jaffe pointed out in their study that no spillover occurs when the citations are added by the patent examiner without the inventor’s awareness. These citations do not have a link with the originating patents, thus should not be more geographically localized than the control patents, and the real spillover in this case is more localized than the citations with “noise”. Moreover, spillovers can occur without citations. This is revealed by the study of Banks (1998) in the study of “NASA”, where he examined evidence of explicit communication between EPB¹³ scientists and the

¹³ Electro-Physics Branch is “involved in the development of high performance durable power materials and surfaces technology to meet NASA, national, and U.S needs.” (Jaffe et al., 1997)

inventors of patents that cite EPB patents.¹⁴ Additionally, basic research cannot be patented, hence there is no citation to basic research.

So far we can conclude that citations are a valid but noisy indicator of spillovers: spillovers can occur without citations and citations can occur without representing any real spillover. Localization of spillovers is not prevalent in all technological categories. For example, in the study of the impact of NASA, Jaffe et al. (1997) found that NASA patents are more localized in “electronic/ nuclear” and less in the other categories. Moreover, in Stolpe (1998)’s study, no strong evidence of localization was found in the field of LCD technology.

¹⁴ In his analysis, only 15% of citations involved known contact, suggesting most spillovers result from indirect rather than direct communication. He also found that for 2 of the 18 patents, contact occurred but without citation. These results are in line with the study of Jaffe et al. (2000): half of all citations did not seem to correspond to any kind of knowledge flow, but one-half of citations correspond to an impact of the cited invention on the citing inventor, and about one-quarter correspond to a fairly rich knowledge flow, a fairly significant impact, or both.

Chapter 4

Using patent citations in market value assessment

Market value is the present discounted value of the future cash flow, which is generated by both tangible and intangible assets. "It is possible and interesting to use the market value of the firm as a partial indicator of the expected success of its inventive efforts." (Griliches, 1981:183)

Much of the previous work on the market value of innovation has been done using patent counts, and they do appear to contain more information than R&D does, as they measure the success of R&D programs (Hall, 1998). But, as Griliches et al. (1987) have shown, patent data are an extremely noisy measure of the value of the patented innovation.

Citation-weighted patents are found to be more informative than patent counts. The breakthrough was made by Trajtenberg (1990), where he suggested using patent citations data instead of simple patent counts to measure the economic value of innovation. In 1997, Shane (1997) regressed Tobin's Q for 11 semiconductor firms using R&D stock, patent stock and patent stock weighted by citations, and found that patents weighted by citations are a more powerful measure than the previous two.

4.1 Market value functions

A popular market value function used in this literature is based on the specification developed by Griliches (1981), relating Tobin's Q to patent stock and

R&D stock. Tobin's Q is a measure of the market value of a firm over its replacement cost, determined both by tangible and intangible assets (Shane, 1997). The model is:

$$V = b(A + K) \quad (4-1)$$

where V is the market value, b is the coefficient of its tangible (A) and intangible (K) assets.

Most of the specifications take a similar form with minor amendments for specific needs. For example, in Hall et al. (2000), the model is:

$$V_{it} = q_t (A_{it} + \gamma_t K_{it})^{\sigma_t} \quad (4-2)$$

where A_{it} denotes the ordinary physical assets of firm i at time t and K_{it} denotes the firm's knowledge assets. Both variables are in nominal terms, γ_t is the shadow value of the knowledge assets. After taking logarithms of both sides, equation (4-2) becomes:

$$\log V_{it} = \log q_t + \sigma_t \log A_{it} + \sigma_t \log(1 + \gamma_t K_{it} / A_{it}) \quad (4-3)$$

The third term is approximated by $\gamma_t K_{it} / A_{it}$, and $\sigma_t \gamma_t$ measures the absolute value of the shadow value of the knowledge assets relative to the tangible assets of the firm (K_{it} / A_{it}). When the model is under constant returns to scale, we have $\sigma_t = 1$. After some manipulations (the log of ordinary physical assets can be moved to left hand side of equation (4-3), and the model can be estimated with Tobin's q as the dependent variable), we obtain the following expression:

$$\log V_{it} / A_{it} = \log Q_{it} = \log q_t + \log(1 + \gamma_t K_{it} / A_{it}) + \delta_t D \quad (4-4)$$

where the dependent variable is Tobin's Q and the last term controls for the overall level of Q when either R&D or patents are missing. The general way to run the

regression is to incorporate R&D stocks, patent stocks, and citation stocks to the knowledge stock or knowledge assets one at a time to compare their performance.

However, most of the studies using patent citations data to measure the value of innovations run into the problem of truncation. This problem occurs when the limit of observation results in a shortage of citations data. The truncation problem can either happen at the beginning or at the end of the period covered by the data. The solution proposed by Hall et al. (2000) is to estimate from the data the shape of the citation-lag distribution. They used a dataset containing over 4800 U.S. manufacturing firms and their patenting activity for the period 1964-1996. However, the citations were only observed starting from 1976. The solution is to assume the distribution to be stationary and independent over time, and estimate the total citations for a given portion of citations which are observed by “dividing the observed citations by the fraction of the population distribution that lies in the interval for which citations were observed” (Hall et al., 2000).

An estimated lag distribution model to adjust the truncation bias is (Hall et al., 2000):

$$D_s = \exp(-\beta_1 s)(1 - \exp(-\beta_2 s)) \quad (4-5)$$

where s is the lag between the application year of the citing patent and the application year of the original patent. The predicted number of citations at time t with lag s can be derived using the following equation:

$$C_{t,s} = D_s \frac{\sum_{j=0}^s C_{t,j}}{\sum_{j=0}^s D_j} \quad (4-6)$$

where t is the application year of a patent, the numerator is the actual number of citations, while D_s and the denominator come from (4-5).

This solution is acceptable when we can observe most of the distribution, but when we only observe a few of the citations in the early years of the patent's life¹⁵, this method will give a very noisy estimation. For this reason, Hall et al. (2000) improved

the model by replacing $\frac{\sum_{j=0}^s C_{t,s}}{\sum_{j=0}^s D_j}$ by the following equation:

$$E \left[\sum_{j=0}^{20} C_{t,j} \mid \sum_{j=0}^s C_{t,j} = 0 \right] \quad (4-7)$$

According to this equation, we can predict the total citations in the 20 years, given the citations that we can observe.

In order to explore the question of whether future citations are relevant to market value analysis, past and future citations can be used in accordance with Hall et al.'s (2000) study. Future citations are the difference between the total citations for a given date and the citations observed by the given date. First, Hall et al. (2000) constructed the "total" citations based on the observed citations at a given date, and then constructed the future citations which are the difference between the total citations and the observed citations. The following equation can be used to estimate the total citations:

$$K^c(t) = E[K^c(t) \mid K^{ppc}(t)] + K^c(t) - E[K^c(t) \mid K^{ppc}(t)] \quad (4-8)$$

¹⁵ e.g. the patent data begins in 1964, but we only observe the citations from 1976, thus the citations granted before 1976 are missing.

where $K^C(t)$ is defined as the total citation weighted patent stock at time t and $K^{PPC}(t)$ denotes the patent stock weighted by the citations received at time t .

Patents weighted by citations and future citations are more significant than the other explanatory variables such as simple patents counts in the analysis of market value. A regression of the log of Tobin's Q on R&D, simple patent counts and citation-weighted patents in Hall et al. (2000) shows obviously that citation-weighted patents do better than patents, since the R^2 (which is one of the criteria to evaluate which independent variables are more correlated to the dependent variable) is higher for citation-weighted patents than for patents, while the coefficients for patents and citations are 0.607 and 0.108, respectively, at a 15% depreciation rate.¹⁶

To find whether citations data add information to the regression, we can look at Table 4-1. Column 2 shows that patents bring to the equation (which already contains an independent variable) only small additional explanatory power. Whereas, after adding average cites per patent to the equation, there is a significant impact on market value, which can be seen from the coefficient of Cite stk/pat stk in column 3, and from the higher R^2 (0.254 compared to 0.222).

As mentioned before, past and future citations can be formed to determine whether or not future citations add information to market value. In table 4-2, we can see from columns 2 and 5 that the coefficient of future citations is always equal to or greater than that of total citations, and that the coefficient of past citations is negative or marginally significant. This implies that future citations are more related to the market value than past citations.

¹⁶ "A firm's specific depreciation rate is based on the rate implicit in financial statements." (Shane, 1997:136) A depreciation rate is used in the estimation since the patent will depreciate in a few years.

Columns 3 and 6 also show that the coefficient of the unpredicted portion of total citations is almost the same as the coefficient of future citations and is higher than that of predicted citations. Hall et al. also found that the market value will be significantly higher for those firms whose patent citations are higher than the median. A survey performed by Harhoff et al. (1997) on citation frequency and the value of patented innovations using U.S and German patents shows that there is a strong relationship between citations and the value of patented innovations. Moreover, patents with greater economic value are more heavily cited in subsequent patents. This is in concordance with what Hall (2000) found in his study of market value, where “firms that average more than the median number of citations per patent have a very significant increase in market value” (Hall et al., 2000:20).

4.2 A non-parametric series estimator to deal with data shortages

Another problem that has been explored is the uncertainty of the market: firms need to make sunk investments relating to the new patent, and they have the option of delaying their investments. When the market conditions are uncertain this will generate real options that measure the value of the option of delaying these investments. To find out whether uncertainty affects market value, the following model (Bloom and Reenen, 2001) can be used:

$$\log\left(\frac{V}{K}\right)_{it} = \delta\left(\frac{G}{K}\right)_{it} + \theta\sigma_i + \zeta\left[\sigma_i * \log\left(\frac{G}{K}\right)_{it}\right] + \tilde{\eta}_i + \tilde{\tau}_t + \tilde{v}_{it} \quad (4-9)$$

where V is the market value of the firm, K is the capital stock, G is the knowledge stock, $\tilde{\eta}$ corresponds to firms' specific fixed effects, $\tilde{\tau}$ is the time effect, and $\tilde{\nu}$ is a random stochastic term. The left side of the equation is the Tobin's average Q .

The truncation problem also appears in the dataset of Bloom and Reenen's study of over 200 major British firms between 1968 and 1996, with 59919 patents in total and citations made to them during the period 1976-1996. To solve the problem, Bloom and Reenen used a non-parametric series estimator based on full Fourier sine and cosine expansion. Four sine and four cosine terms for the Fourier expansion were used to fit a smooth curve to the observed time variation in aggregate citation levels to non-parametrically estimate a truncation bias function, and then the normalized citations for each patent remain constant over the period. Figure 4-1 shows the actual and normalized average citations per year. Bloom and Reenen constructed three measures: patent stock, normalized citation weighted patent stock and a five year citation stock from 1990 until 1996 without truncation bias correction. To avoid the problem of truncation bias, only the data up to 1993 for patents and 1990 for citations are used.

The regression results comparing the effect of simple patent counts and citation weighted patents on Tobin's Q are shown in table 4-3. The coefficients are 1.620 and 0.427 for the patent stock and the cites stock, respectively. A depreciation rate of 30% was used. This result is higher than the coefficients for patents and citations derived in Hall et al.'s (2000) study, which are 0.607 and 0.108, respectively, with a 15% depreciation rate.

4.3 Using expected weighted patent counts to eliminate the truncation bias

The same truncation problem appeared in Shane's study of the semiconductor industry. Shane used data on patents and patent citations from 1972-1992 for the semiconductor industry. The expected weighted patent count (EWPC) is used to eliminate the truncation bias. EWPC equals the sum of the actual citations received until 1992 plus the expected citations which would be received. A regression model generates the EWPC is presented in Shane (1997)'s study:

$$C_T = \alpha_T + \sum_{t=0}^{T-1} \beta_{tT} C_t + \varepsilon \quad \text{for } T = 2, 3, \dots, 17 \quad (4-10)$$

$$EWPC = \sum_{t=0}^{(1992-year)} C_t + \sum_{t=1992-year+1}^{17} \hat{C}_t \quad (4-11)$$

For example, when T=2, the citations in the second year can be predicted by estimating coefficients based on the citations received in the issuance year and the first year, when T=17, model (4-11) can be used to predict citations received in year 17 based on citations in the issuance and the subsequent 16 years.

In Shane's study, a quadratic time trend was added to the model, and the performance of simple patent counts stock (SPS), R&D and EWPS are compared. These models are shown below:

$$\log(q_{f,t}) = \phi + \mu_1 t + \mu_2 t^2 + \alpha_{SPS} (SPS_{f,t} / A_{f,t}) + \alpha_{R\&DS} (R \& DS_{f,t} / A_{f,t}) + \varepsilon_{f,t} \quad (4-12)$$

$$\log(q_{f,t}) = \phi + \sum \mu_i D_i + \alpha_{SPS} (SPS_{f,t} / A_{f,t}) + \alpha_{R\&DS} (R \& DS_{f,t} / A_{f,t}) + \varepsilon_{f,t} \quad (4-13)$$

$$\log(q_f) = \phi + \mu_1 t + \mu_2 t^2 + \alpha_{EWPS} (EWPS_f / A_f) + \alpha_{R\&DS} (R \& DS_f / A_f) + \varepsilon_{f,t} \quad (4-14)$$

$$\log(q_f) = \phi + \sum \mu_i D_i + \alpha_{EWPS} (EWPS_f / A_f) + \alpha_{R\&DS} (R \& DS_f / A_f) + \varepsilon_{f,t} \quad (4-15)$$

Equations (4-12) and (4-14) were used to compare the performance of simple patent counts and expected weighted patent counts with R&D. The variable $q_{f,t}$ is the market value measured by Tobin's Q for firm f at time t , $A_{f,t}$ is defined as the tangible assets of firm f at time t , $SPS_{f,t}$ and $R \& DS_{f,t}$ were constructed as firm f 's simple patent stock and R&D stock at time t , $EWPS_f$ is constructed in the same way as $SPS_{f,t}$ based on the future citations for firm f . Alternatively, equations (4-13) and (4-15) were augmented with D_i as a dummy variable for year i to control for the time effect.

In Shane's study, the regression analysis shows that $EWPS$ has a significant coefficient with a positive sign, while WPS has a negative sign and is not significant. This strongly suggests that future citations are more correlated with market value instead of simple patent counts.

4.4 Summary

The studies reviewed in this section come to a common finding: citations provide a better explanation of the market value of the firm than simple patent counts.

Moreover, more recent studies show that market value is more correlated with future citations. It is this part of citations which generates more information about future value.

In terms of the methods used, a dominant model has been adopted from Zvi Grilches and has been developed. One important addition to the model has been the time effect. Without the time effect, the model gives poor results because of changes in the tax system and the business cycle.

The main problem with the data remains the truncation bias which cannot be easily avoided. Because patent value has more to do with predicted or future citations than with past citations, it is important to predict future citations precisely, thus the correction of the truncation bias is crucial.

Chapter 5

Using patent citations in innovation value assessment

Innovations vary enormously in value: some innovations are important, triggering later research in the relevant field, while others are trivial. In a growing literature, researchers have put a lot of effort in measuring the value and importance of innovations. Patent citations have proved to be a good indicator among the leading indicators of the ex post value of innovations.

Simple patent counts are not a very good indicator of the value of innovations, because they are indicative only of the input side of innovation as reflected by research and development. Simple patent counts simply assume that all patents have the same value, ignoring skewness in the value of patents. Patent citations thus became widely used as an indicator of the ex post value of innovation instead of simple patent counts.

5.1 Citations and the value of innovation

Patent citations have been compared with simple patent counts (SPC) and other explanatory variables by looking at the regression results examining how patent citations are related to innovation. In Trajtenberg (1990), the following model has been used to calculate the social value of innovation between periods t and $t-1$:

$$\Delta W_t = W(S_t) - W(S_{t-1}) \quad (5-1)$$

where $W(\cdot)$ stands for the social surplus function, and where S_t and S_{t-1} are the sets of products provided in two successive periods, and ΔW_t represents the value of innovation. The total gain associated with the innovations at time t can be expressed as:

$$TW_t = \Delta W_t \left[n_t + K \left(\sum_{\tau=0}^t \Delta W_\tau \right) \int_{t+1}^{\infty} f(\tau) e^{-r(r-t-1)} d\tau \right] \equiv \Delta W_t (n_t + n_f) \quad (5-2)$$

where n_t is the number of consumers at t , $K(\cdot)$ is the ceiling of the diffusion curve, which is defined as “the maximum number of individuals that will end up adopting the innovation in the long run, expressed as a percentage of the population of potential adopters.” In Trajtenberg (1990), $f(\cdot)$ is a diffusion path that “traces the distribution of some sort of reservation price in the population of potential adopters.” (Trajtenberg, 1990:158) and r is the interest rate, TW_t is the total gain at time t . With these two equations, the relationship between simple patent counts and weighted patent counts and ΔW and TW can be tested.

The dataset used for CT¹⁷ scanner is from 1971 to 1986, 456 patents in total. Patents are dated according to the application date rather than the granting date. Griliches et al. (1987) found that patent counts dated by application date are more tightly linked to the economic value of the innovation than the patent counts dated by the granting date. The dataset covered the period 1972-1982, and the citation data was from 1983 to 1986. Since there is no attempt to explore spillovers, only the citations “within referencing”, which belong to the same field are counted. This contrasts with Shane (1997)’s study of market value, where all citations were used, including those not belonging to the semiconductor technology. During the process of constructing the

¹⁷ Computed Tomography.

dataset, the problem of age and truncation bias was also explored, and the hypothesis that older patents received more citations just because of the passage of time was rejected. The truncation bias was proved too weak to affect the results, and hence was ignored.

Table 5-1 and figure 5-1 give us a first look at the comparison between simple patent counts and weighted patent counts based on citations. According to this figure, simple patent counts peak in 1977, and then drop gradually, while the counts weighted by citations shift the graph back in the early years. Comparing the two distributions, the difference of the means is around seventeen months, the weighted patent count centers around mid-1976, and there is a clear trend showing that earlier patents received more cites than the later ones. *WPC* drops from 73 in 1972 to 6 in 1984.

Table 5-2 shows the correlation between patent count and the value of innovation represented by ΔW and TW , both for all the patents and for the patents granted in the CT field only. The most interesting finding is that *WPC* is more correlated with the value of innovation than *SPC*. This correlation increased when only the patents to firms in CT were considered, which implies that there are also patents granted to the assignees in other fields besides CT manufactures, and “the appropriability of patents granted to other assignees was not nil.”(Trajtenberg, 1990:181)

Trajtenberg carried out a test to prove that there is a strong correlation between *SPC* and R&D, and a much weaker correlation between R&D and *WPC*, suggesting simple patent counts are good indicators of the input side of innovation as measured by R&D. Moreover, the correlation is marginally higher for the *SPC* of all patents than for

patents of firms in CT, suggesting a possible spillover from CT manufactures to other fields.

5.2 Citations and innovation quality with a latent model

Citations can also be used in a latent model in order to exploit information embedded in the characteristics of patents, to make an early assessment of innovation quality. Consider the following model, due to Lanjouw and Schankerman (1999):

$$y_{ki} = \mu_k + \lambda_k q_i + e_{ki} \quad q \sim N(0,1) \text{ and } e_k \sim N(0, \sigma_k^2) \quad (5-3)$$

where y is defined as the observed vector of indicators, k denotes the indicators, these indicators include: the number of backward and forward citations, family size and claims, i denotes the patent, q is the common factor quality¹⁸, and represents the technological significance of an innovation. λ_k is the coefficient for indicator k . Among the indicators, *Claims* delineates what is protected by the patent, “the principal claims define the essential novelty of the invention and the subordinate claims detail the feature of the innovation”(Lanjouw and Schankerman, 1999:3), *Family size* is the group of patents protecting the same innovation constitutes its “family”, *forward citations* and *backward citations* are used representing the number of prior patents cited in the application and subsequent patents that had cited a given patent in their own application, respectively. This one-factor latent variable model tells us which indicator is more correlated with the quality of innovation. In addition, equations 5-4 and 5-5

¹⁸ Common factor is the unobserved characteristic of a patented innovation that influences all of the indicators.

show the relative contribution of each indicator to the “composite index” of quality, and also the variance reduction conditional on different set of indicators.

$$E[q | y] = \lambda' \Lambda^{-1} y \quad (5-4)$$

$$Var(q | y) = 1 - \lambda' \Lambda^{-1} \lambda \quad (5-5)$$

In some other studies (e.g. Trajtenberg, 2000 and Harhoff et al., 1997), some simple OLS regressions are carried out using citations as the dependent variable and the social value of innovation as an explanatory variable, expecting the value to be positively related to citations.

The data used in Lanjouw and Schankerman (1999)’s study comprises U.S. patents applied for during the period 1960-1991, and are divided into four categories: pharmaceuticals, chemicals, electronic and mechanical. For this study, truncation was used on each variable.

Table 5-3 shows the results using multiple indicators with the one-factor model that was mentioned at the beginning of this section. The point estimate of λ_{fwd} is greater than the point estimate for the other indicators, suggesting that forward citations are more correlated with the quality of innovation, and also implying that an innovation with many citations is more valuable than an innovation with less citations. The table also presents the signal ratio denoted by S_k which is defined as the percentage of the variance of the k_{it} indicator associated with the common factor q , the value of innovation. When $\sigma_v^2 = 1$, $S_k = \frac{\lambda_k^2}{\sigma_{y,k}^2}$, where $\sigma_{y,k}^2$ is the variance of the k_{it} indicator demeaned by nationality and technology group. The signal ratio for forward citations is larger in drugs and chemicals than in other fields.

The weights for each indicator are derived using the parameters estimated from table 5-4. The weight that forward citations get is about around 1/3, claims and backward citations get 1/4 each, and family size gets the remainder. Furthermore, the conditional variance of quality varies when using equation (5-5) and different subsets of the indicators to predict the latent variable, and we observe a big variance reduction caused by the forward citations as shown in table 5-4. The unconditional variance is normalized to unity as shown in the first row, and each entry represents the equation (5-5): $\text{Var}(q | y) = 1 - \lambda' \Lambda^{-1} \lambda$. The second row is the condition on all four indicators, reducing the variance by about two-thirds.

We can tell from the figures in the table that forward citations constitute the most important indicator. When only Fwd5¹⁹ was dropped, the variance increased by 43% for the pooled sample, and by 75% for the drugs group ($0.58 - 0.329 / 0.329 = 75\%$). The effect is not so severe when the other indicators are dropped. When dropping only claims, the variance increases by 35% for the pooled sample, but this difference varies among the fields. The variance is less for the fields of drugs and chemical patents (around 14% to 20%) than for electronic and mechanical patents (around 40% to 50%). The authors also explored whether citations with a longer span are more helpful in measuring the initial expectations about the quality of patented innovations, and found that later citations are less correlated with the quality of the innovation, implying later citations are distantly related to the innovation.

¹⁹ Fwd5 includes forward citations to the patent that occur within five years of the patent application date (Lanjouw and Schankerman, 1999)

These results obtained with multiple indicators strengthen the belief that forward citations are the least noisy indicator compared with the results obtained using other indicators.

5.3 Summary

A common finding among the studies reviewed in this section is that patent counts weighted by citations (forward citations) are the most relevant indicator of the value of innovation. Other indicators, such as simple patent counts, backward citations, and patent family size fail to capture the value or quality of innovation.

Chapter 6

Patent citations in other fields

In addition to the use of citations in the study of knowledge spillovers, market value and innovation value, patent citations are used in a few other areas.

First, patent citations have a significant impact on firm level productivity; this has been studied by Bloom and Reenen (2001) using a database on over 200 major British companies between 1968 and 1996, covering 59,919 patents. In his findings, the citation weighted patent stock is significant at the 0.05 level with an elasticity of 0.03, suggesting that doubling the patent stock would increase productivity by 3%. However the result is not significant when both patents and citations are added as explanatory variables, suggesting citations indeed provide some useful information.

Second, patent citations are found to have a strong impact on patent renewal behavior. Using multiple indicators to study innovation, Lanjouw and Schankerman (1999) found that forward citations and family size are important for renewal decisions, but not the other indicators. Maurseth (2001), when using a dataset including patents applied for by Norwaegians and patents applied for in the European Patent Office (EPO), finds that a cited patent with a higher value has a longer survival period and will be cited more frequently. On the other hand, when the citations occur mainly within the field, patents receiving citations are renewed for shorter periods.

Finally, there are some interactions between patent citations and firms' size as suggested by Almeida and Kogut (1996). They used two types of data: patents relating to important semiconductor design innovations, and semiconductor plant location data

used to identify relevant regions. The most important regions that have been studied are the Silicon Valley area and the New-York/New-Jersey/Pennsylvania area. Based on these datasets, two samples of major patents including start-ups and large firms have been analyzed by constructing control samples. The results suggest that knowledge seems more localized in start-up companies, and that the latter produce innovations in less crowded fields, while large firms concentrate in the more established fields. Moreover, start-ups are more tightly related to knowledge networks than the larger firms are.

Chapter 7

Avenues for future research

7.1 Patent citations made by the examiner and self-citations

As explained in chapter 2, both the applicant and the patent examiner make the patent citation list. Thus citations can be divided into two types, those made by the applicant and those made by the examiner. The distinction does not matter when patent citations are used in analyzing firm market value or innovation value; these variables have nothing to do with who made the citation.

Whereas, when citation data is used in studying knowledge spillovers, the citation type matters. Citations made by the applicant represent knowledge spillovers, while citations made by the examiner do not. Thus, citations added by the patent examiner will introduce considerable noise in the results. The only case where examiner-added citations represent spillovers is when the applicant purposely conceals a due reference to a previous patent. But this is impossible to verify, both for the patent examiner and for the researcher.

The same logic can be extended to self-citations. Self-citation is simply a citation assigned to the same organization as the originating patent. Self-citation is generally regarded as not involving any knowledge spillover.

Failing to consider these two issues leads to imprecise conclusions about knowledge spillovers. Generally speaking, citations made by the examiner will favor a conclusion of non-localization of knowledge spillovers, whereas self-citations will

support a conclusion of localization of knowledge spillover. Hence, the biases are in opposite directions.

In most studies, self-citations are weeded out during the preprocessing of patent citations data. It is much more difficult to control for the citations made by the examiner. Usually, economists simply neglect the existence of citations made by the latter. Unfortunately, there is no study validating such neglect and we do not know to what extent such neglect biases the results.

7.2 Truncation of the patent citation data

Due to the nature of the patent citation data, there is always a truncation problem. The truncation problem occurs because the time period of patent data based on the grant date does not always match that of the citations data, resulting in missing citations. Truncation of the data at the beginning is usually the result of unavailability of the citation data, while truncation at the end is due to the fact that more recent patents have had less time to receive citations, or that the sample period does not extend to the present. For instance, in Hall et al. (2000), the available patent data is from 1964 to 1996 and the available citation data starts from 1976. This results in truncations at both the beginning and the end. For example, a 1964 patent is likely to be cited before 1976 but the citation data is not available until 1976, which leads to truncation at the beginning. A 1993 patent is likely to be cited after 1996 but the citations data is only available before 1996 which results in truncation at the end.

Different approaches have been taken to resolve this issue. One is estimating the total citations from an observable portion of citations by “dividing the observed citations by the fraction of the population distribution that lies in the interval for which citations were observed” (Hall, et al., 2000:10). This method works well when we know most of the distribution of citations. It is problematic when we know only a small portion of that distribution.

Shane (1997) used a regression model to estimate the expected weighted patent counts based on the observed patent citations, and Bloom and Reenen (2001) used a non-parametric series estimator based on full Fourier sine and cosine expansion. Their methods are based on the assumption that truncation leads to biased results. However, not all of the studies support this point. Trajtenberg (1990) used a series of statistics to prove that a truncation bias does exist and that a small number of missing citations are not sufficient to bias the results. However, Trajtenberg does not use future citations, which may well explain the smaller bias he estimates. In all cases, truncation is likely to remain a major issue in future studies of patent citations.

7.3 The value of the cited patent

Regarding the value of innovation, most studies suggest that the more citations to a patent occur, the more valuable the innovation is, and hence the cited patent will be more useful in later research. Research on both firm market value and innovation value confirms this point. However, this view has recently been challenged.

Maurseth (2001) found that patents cited across fields are renewed for longer periods and imply a higher economic value. Patent citations within the same fields have an opposite effect: these patents are renewed for a shorter period than the patents cited across fields. Citations within a field indicate an increase in the rival's value of innovation, which means the more citations coming from the same field, the lower is the value of the cited patent. Based on this reasoning, citations coming from the same field may not cause a problem when analyzing knowledge spillovers, but may constitute an issue when dealing with market value and innovation, as they are reflective of a dilution in the value of the cited patent.

Chapter 8

Conclusions

Patent citations carry important information relating to patent value that cannot be captured by simple patent count and other proxies. In this paper, we did a detailed review of major fields of research where patent citations data has been commonly used.

Patent citations can be used in the study of knowledge spillovers. It is found that spillovers can flow across countries, and across different units and organizations. In terms of our review of the localization of spillovers, most of the studies concluded that spillovers are localized among different technological categories to different extents; on the other hand, some fields are characterized by non-localization (e.g. LCD technology). Nevertheless, citations data are also shown to be a noisy indicator (e.g. Jaffe et al., 1993 and Jaffe et al., 2000). Quite a number of citations are not added by inventors themselves, instead, they are added by patent examiners. These citations do not carry any knowledge spillovers. Moreover, spillovers sometimes occur without citations; for example, basic research spills over without citations.

Patent citations are also used in the study of market value. Citations are found to be more correlated with market value than other indicators. Furthermore, future citations are more highly correlated with market value than simple patent counts.

In terms of using citation data in estimating the quality of innovation, there is strong evidence that, when compared with other indicators, patent citations have a stronger correlation with innovation.

Appendix

Table 3-1

DESCRIPTIVE STATISTICS						
Oraginating Dataset	Percent Receiving Citations	Total No. of Citations	Mean Citations Received	Average Citation Lag	Percent Self Citaions	Percent Same Patent Class
1975						
University	88.6	1933	6.12	6.53	5.6	54.3
Top Corporate	84.2	1476	4.7	7.17	18.6	55.7
Other Corporate	82.3	1341	4.22	7.82	9.1	57.5
1980						
University	79.9	2093	4.34	4.36	8.9	56.3
Top Corporate	79.9	1701	3.54	4.41	24.6	58.3
Other Corporate	74.1	1424	2.95	4.46	12.6	57.2

Source: Jaffe et al. (1993)

Table 3-2

GEOGRAPHIC MATCHING FRACTIONS						
	Originating Cohorts					
	University	1975		University	1980	
		Top Corporate	Other Corporate		Top Corporate	Other Corporate
Number of Citations	1759	1235	1050	2046	1614	1210
<u>Matching by Country</u>						
Overall Citation Matching Percentage Citations	68.3	68.7	71.7	71.4	74.6	73
Excluding Self-Cites	66.5	62.9	69.5	69.3	68.9	70.4
Controls	62.8	63.1	66.3	58.5	60	59.6
t-statistic	2.28	-0.1	1.61	7.24	5.31	5.59
<u>Matching by States</u>						
Overall Citation Matching Percentage Citations	10.4	18.9	15.4	16.3	27.3	18.4
Excluding Self-Cites	6	6.8	10.7	10.5	13.6	11.3
Controls	2.9	6.8	6.4	4.1	7	5.2
t-statistic	4.55	0.09	3.5	7.9	6.28	5.51
<u>Matching by SMSA</u>						
Overall Citation Matching Percentage Citations	8.6	16.9	13.3	12.6	21.9	14.3
Excluding Self-Cites	4.3	4.5	8.7	6.9	8.8	7
Controls	1	1.3	1.2	1.1	3.6	2.3
t-statistic	6.43	4.8	8.24	9.57	6.28	5.52

Source: Jaffe et al. (1993)

Table 3-3

Weighted Probit Model of Patent Citations-Estimated Coefficients (Standard Errors in Parentheses)

Citing country	All countries			US	Japan
Constant	-24.609 (1.941)	-24.551 (1.029)	-24.562 (1.023)	-23.717 (2.169)	-25.926 (1.287)
Appropriability	-0.57 (0.302)	- -	- -	-0.543 (0.596)	-0.312 (0.365)
Basicness	0.321 (0.063)	0.315 (0.047)	0.315 (0.046)	0.298 (0.115)	0.368 (0.077)
Subclass Equality	12.356 (1.198)	12.312 (1.188)	12.304 (1.187)	7.324 (1.959)	19.592 (1.701)
Examiner Equality	0.275 (1.555)	- -	- -	- -	- -
Assignee Equality	0.831 (2.568)	0.846 (2.458)	- -	6.847 (3.701)	-0.289 (2.150)
Prior Cooperation	-0.962 (3.024)	-0.918 (2.940)	-0.26 (2.043)	- -	- -
Large Inventor Distance	-0.065 (1.575)	- -	- -	- -	- -
Foreign Inventor Dummy	-0.105 (1.571)	- -	- -	- -	- -
Lag 4 to 6 years	0.182 (1.418)	0.061 (1.243)	0.097 (1.240)	0.0487 (2.409)	-0.38 (1.573)
Lag 7 to 9 years	0.242 (1.447)	0.11 (1.203)	0.134 (1.203)	0.425 (2.251)	-0.158 (1.561)
Lag over 9 years	-0.084 (1.541)	-0.226 (1.307)	-0.207 (1.306)	-0.469 (2.482)	-0.283 (1.692)

Source: Extracted from Stolpe (1998)

Table 3-4

Summary of Geographic Localization Effects Based on Citations to Patents Originating in the "NASA Complex" of California, Texas, Ohio, Alabama and Virginia/Maryland/DC

	Fraction of citations coming from Same State		Fraction of citations coming from Other "Complex" States		Fraction of citations coming from All Other U.S.		Fraction of Foreign Citations	
Raw Population Means:								
All Patents	0.184		0.121		0.417		0.278	
NASA Patents	0.128		0.188		0.393		0.292	
Other Federal Patents	0.158		0.177		0.379		0.287	
Regression Results:								
	Standard		Standard		Standard		Standard	
	Coefficient	Error	Coefficient	Error	Coefficient	Error	Coefficient	Error
NASA Patents	-0.036	0.008	0.033	0.008	-0.004	0.0103	0.007	0.0096
Other Federal Patents	-0.019	0.0063	0.025	0.0063	-0.007	0.0082	0	0.0076
California	0.104	0.0084	-0.119	0.0085	0.01	0.011	0.005	0.0102
Texas	0.05	0.0119	-0.052	0.012	0.015	0.0155	-0.013	0.0145
Virginia/DC/Maryland	0.02	0.0084	-0.031	0.0085	-0.003	0.011	0.014	0.0102
Ohio	0.033	0.01	-0.054	0.01	0.017	0.013	0.003	0.0121
R-squared	0.096		0.066		0.049		0.046	
Dependent Mean	0.158		0.17		0.387		0.286	
Root M.S.E	0.268		0.27		0.349		0.326	

Source: Jaffe et al. (1997)

Table 3-5

Summary of Geographic Localization Effects Based on Citations to Patents Originating in the "NASA Complex" of California, Texas, Ohio, Alabama, and Virginia/Maryland/DC (Excluding Self-Cited)

	Fraction of citations coming from Same State	Fraction of citations coming from Other "Complex" States	Fraction of citations coming from All Other U.S.	Fraction of Foreign Citations				
	Raw Population Means:							
All Patents	0.127	0.124	0.428	0.322				
NASA Patents	0.079	0.18	0.423	0.318				
Other Federal Patents	0.104	0.182	0.369	0.317				
	Regression Results:							
	Standard Coefficient	Standard Error	Standard Coefficient	Standard Error	Standard Coefficient	Standard Error	Standard Coefficient	Standard Error
NASA Patents	-0.019	0.0068	0.02	0.0085	0.009	0.011	-0.01	0.0105
Other Federal Patents	-0.013	0.0064	0.026	0.0067	-0.002	0.0087	-0.011	0.0083
California	0.129	0.0072	-0.131	0.009	0.002	0.0116	0.001	0.0111
Texas	0.072	0.0103	-0.069	0.0127	0.014	0.0165	-0.017	0.0157
Virginia/DC/Maryland	0.042	0.0086	-0.06	0.0106	0.02	0.0137	-0.001	0.0131
Ohio	0.042	0.0086	-0.06	0.0106	0.02	0.0137	-0.001	0.0131
R-squared	0.087		0.066		0.045		0.046	
Dependent Mean	0.104		0.173		0.404		0.318	
Root M.S.E	0.224		0.278		0.36		0.343	

Source: Jaffe et al. (1997)

Table 4-1

Effect of Adding Patents and Citations to R&D Regression
 U.S. Manufacturing Firms (Cleaned Sample)-1979-88-12,119 firm-years-1,983 Firms
 Nonlinear Model with Dependent Variable=logarithm of Tobin's q

Independent Variable	(1)K	(2) K with P/K	(3) P/K and C/P	(4)K with C/K
R&D Stock(K)/Assets	1.198(0.56)	1.276(0.061)	1.366(0.076)	1.296(0.061)
D(R&D=0)	0.015(0.018)	0.058(0.019)	0.0666(0.019)	0.095(0.019)
Pat Stock/K		0.0268(0.0062)	0.0306(0.0076)	
Cite Stk/Pat Stk			0.0514(0.0039)	
Cite Stock/K				0.0092(0.0013)
R-squared	0.218	0.222	0.254	0.231
Standard error	0.687	0.685	0.671	0.681

Source: Hall et al. (2000)

Table 4-2

Splitting Total Citation Stock into Past and Future
 U.S. Manufacturing Firms (Cleaned Sample) - 1979-88-12,119firm-years-1,983 Firms
 Nonlinear Model with Dependent Variable=logarithm of Tobin's q

Independent Variable	Cite/Assets		P/K and C/P	
	(2)	(3)	(5)	(6)
R&D Stock(K)/Assets			1.267(0.07)	1.136(0.059)
D(K=0)			0.063(0.018)	0.06(0.018)
Cite Stock/A				
Past Cite Stk/A	-0.0554(0.0166)			
Future Cite Stk/A	0.1437(0.0073)			
Pred.Cite Stk/A		0.1140(0.0054)		
Unpred.Cite Stk/A		0.077(0.0076)		
Pat Stock/K			0.0284(0.007)	0.0225(0.0057)
Cite Stk/Pat Stk				
Past Cite Stk/P Stk			-0.0156(0.0068)	
Future Cite Stk/P Stk			0.0601(0.0046)	
Pred.Cite Stk/P Stk				0.0118(0.0037)
Unpred.Cite Stk/P Stk				0.05(0.003)
R-squared	0.212	0.207	0.26	0.26
Standard error	0.69	0.693	0.668	0.669

Source: Extracted from Hall et al. (2000)

Table 4-3

Market Value with Patent Measures

$\log(V_{i,t}/K_{i,t-1})$	(1)	(2)	(3)	(4)
Patent Stock/capital	1.62 (0.537)			-0.352 (0.828)
Cites Stock/Capital		0.427 (0.147)		
5 Year Cites Stock/Capital			0.519 (0.221)	0.491 (0.243)
Firm Dummies	yes	yes	yes	yes
Time Dummies	yes	yes	yes	yes
No.Observations	2053	1748	1748	1748
No.Firms	205	182	182	182

Source: Bloom and Reenen (2001)

Table 5-1

Patents in Computed Tomography: Counts and Citations by Year

Year	patents	
	simple patent counts(spc)	weighted by citations(wpc)
1972	1	73
1973	3	50
1974	21	199
1975	48	242
1976	66	235
1977	115	260
1978	71	126
1979	59	88
1980	26	33
1981	15	18
1982	12	13
1983	13	14
1984	6	6
All	456	1357

Source: Trajtenberg (1990)

Table 5-2

Correlations' of Simple and Weighted Patent Counts

Lag	All Patents		Patents to firms in CT	
	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	0.48 (0.16)	0.6 (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	-0.09	-0.175	0.027

Source: Trajtenberg (1990)

Table 5-3Parameter Estimates for One-Factor Model, By Technology Groups^a

Independent Variable(log)	Drugs	Chemical	Electronic	Mechanical	Pooled
Fwd5	0.68 (0.07)	0.61 (0.08)	0.55 (0.04)	0.44 (0.03)	0.52 (0.02)
Claims	0.38 (0.06)	0.41 (0.06)	0.45 (0.04)	0.46 (0.03)	0.44 (0.02)
Family	0.35 (0.06)	0.32 (0.06)	0.39 (0.04)	0.38 (0.03)	0.37 (0.02)
Bwd Cites	0.4 (0.06)	0.31 (0.06)	0.24 (0.04)	0.27 (0.03)	0.28 (0.02)
S _{fwd}	0.46 (0.09)	0.37 (0.09)	0.3 (0.05)	0.19 (0.03)	0.27 (0.02)
S _{clm}	0.14 (0.06)	0.17 (0.06)	0.2 (0.04)	0.21 (0.03)	0.19 (0.02)
S _{fam}	0.13 (0.06)	0.1 ^b (0.06)	0.15 (0.04)	0.14 (0.03)	0.14 (0.02)
S _{bwd}	0.16 (0.06)	0.1 ^b (0.06)	0.06 (0.03)	0.07 (0.03)	0.08 (0.02)
No. obs.	615	606	1767	3123	6111
$\chi^2(2)$	0.39	0.79	5.2	2.86	8.04

Note:

- a) Estimated standard errors are in parentheses. The model also includes nationality dummy variables (and for the pooled data, technology dummies). An 'S_k' denotes the estimated signal ratio calculated as $\lambda_k^2 / \sigma_{y,k}^2$. The $\chi^2(2)$ statistic tests the over-identifying restrictions for the one-factor model.
- b) Not significant at the 0.05 level.

Source: Lanjouw and Schankerman (1999)

Table 5-4Conditional Variance of the Latent Variable: Different Sets of Indicators^a

Indicators(log)	Drugs	Chemicals	Electronics	Mechanical	Pooled
Unconditional Variance	1	1	1	1	1
Fwd5, Claims, Family, Bwd	0.329	0.387	0.36	0.337	0.361
Drop Fwd5	0.58	0.626	0.502	0.46	0.518
Drop Claims	0.378	0.472	0.495	0.498	0.489
Drop Family	0.341	0.407	0.43	0.406	0.412
Drop Fwd5 and Family	0.625	0.673	0.622	0.56	0.606
No. obs.	4721	3450	5234	2656	16061

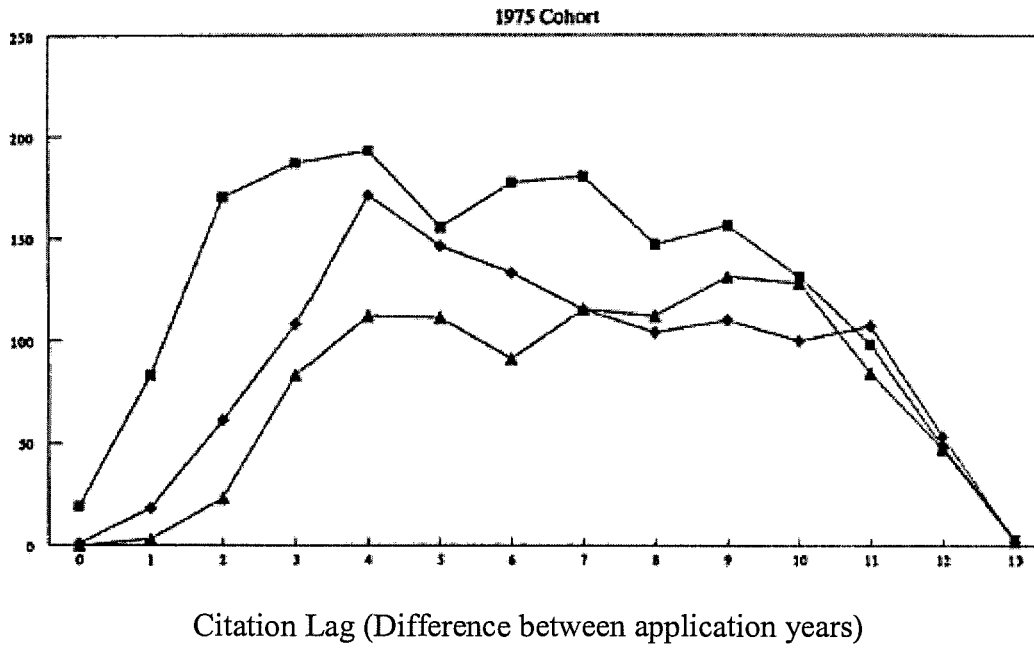
Note:

a) Computed as the predicted value of $Var(q | y) = 1 - \lambda\Lambda^{-1}\lambda$ using estimates of the factor loadings, λ , and covariance matrix Λ for the set of indicators relevant in each case.

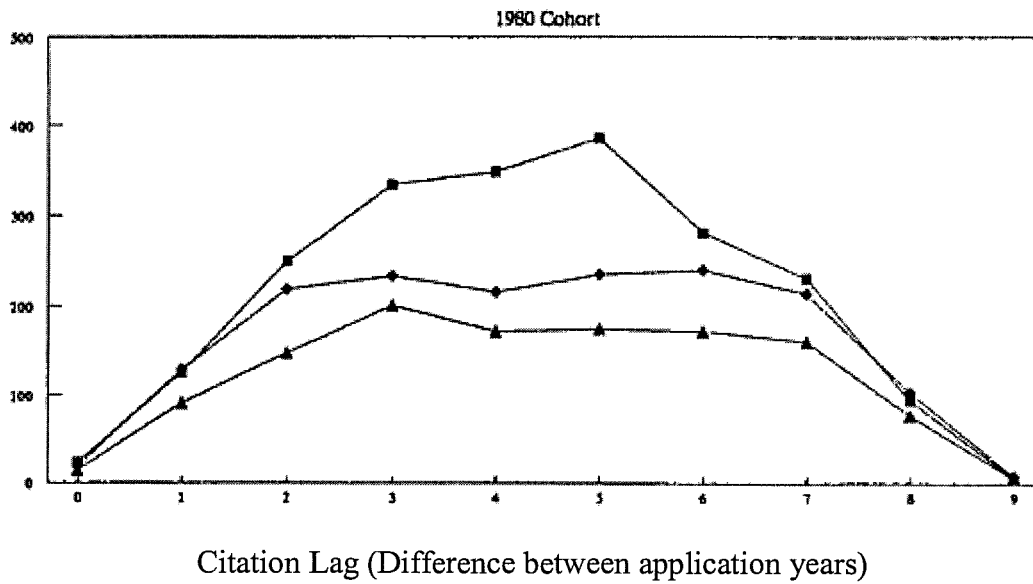
Source: Lanjouw and Schankerman (1999)

Figure 3-1

Number of citations:



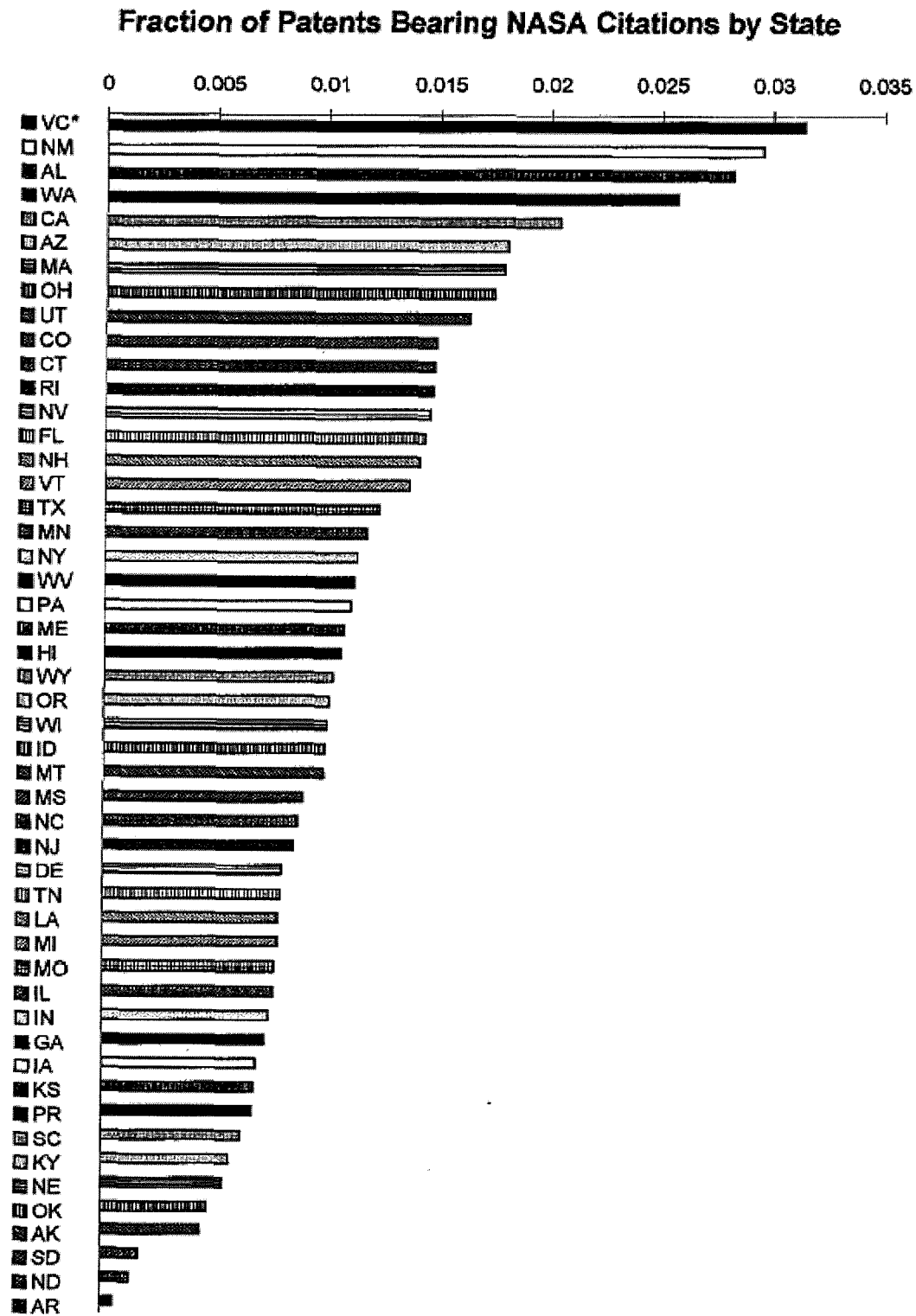
Number of citations:



- university citations
 - ▲ other cooperate citations
- ◆ top cooperate citations

Source: Jaffe et al. (1993)

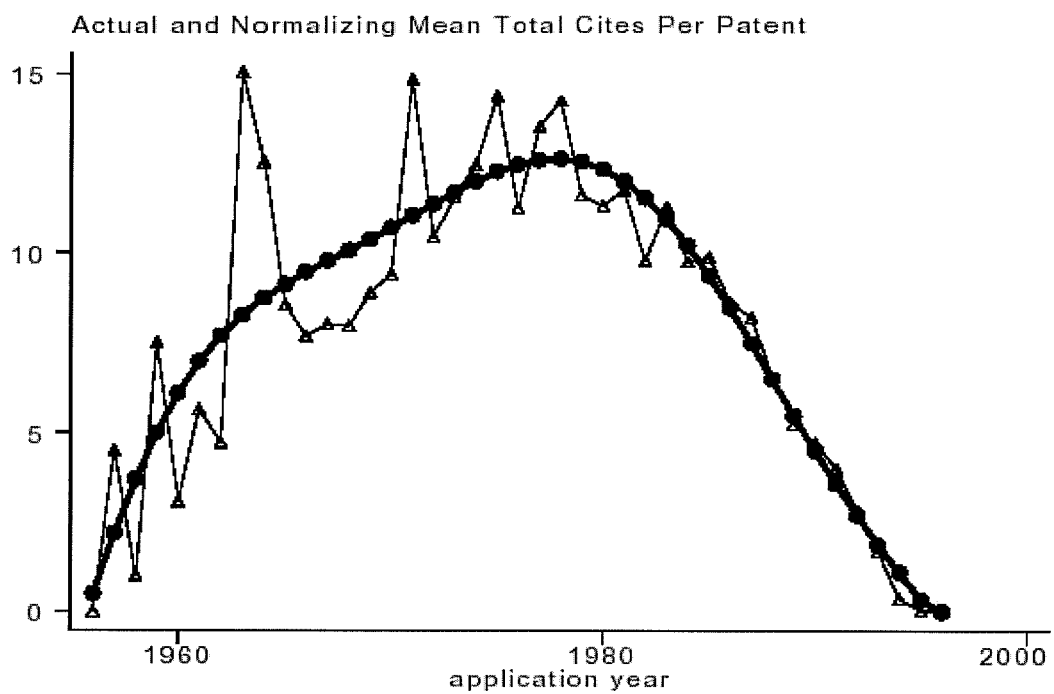
Figure 3-2



*Virginia/Maryland/DC

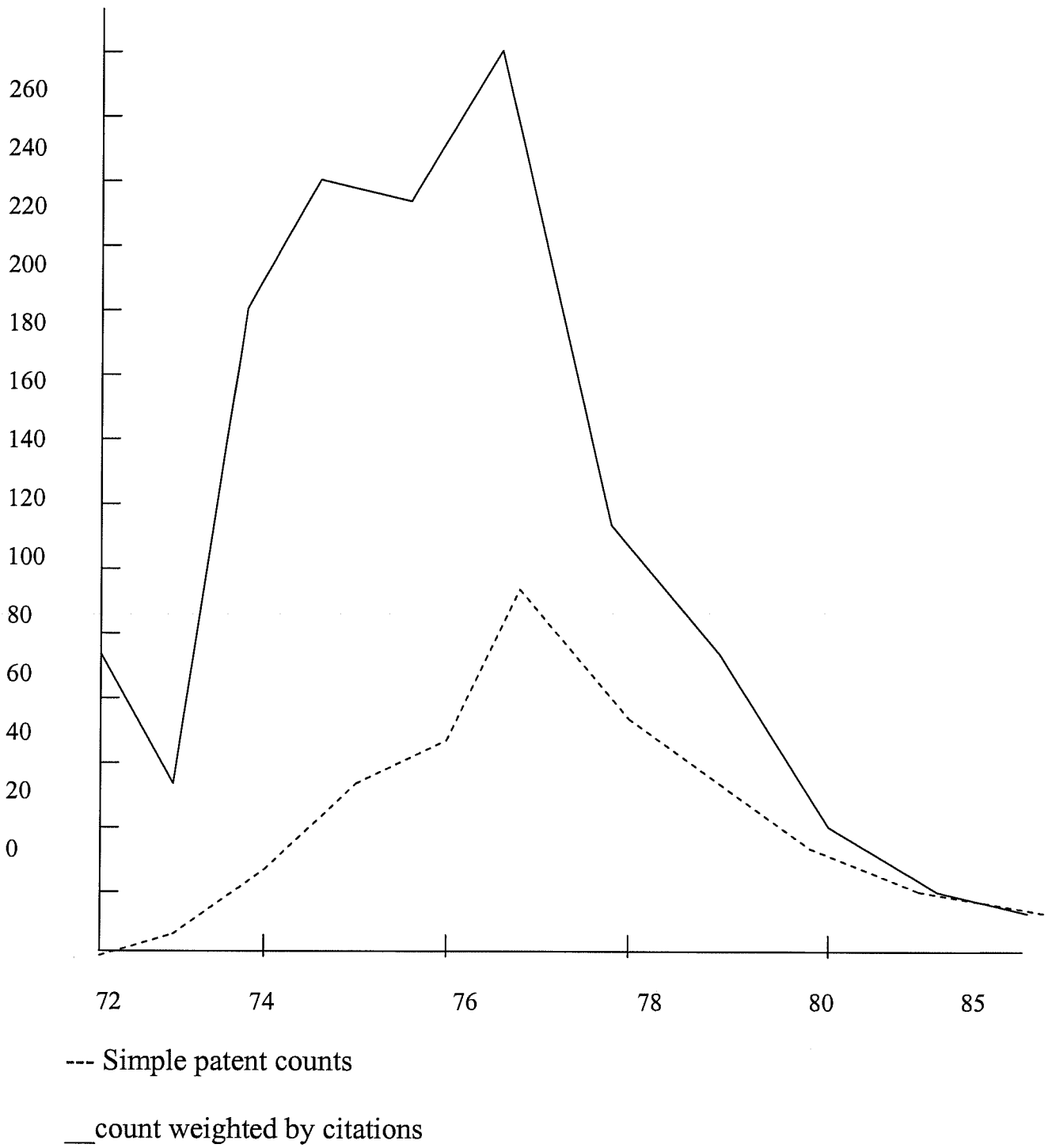
Source: Jaffe et al. (1997)

Figure 4-1



Source: Bloom and Reenen (2001)

Figure 5-1



Source: Jaffe et al. (1993)

Bibliography

- Almeida, Paul and Bruce, Kogut (1997) "The Exploration of Technological Diversity and the Geographic Localization of Innovation", *Small Business Economics*, vol 9, 21-31
- Bloom, Nicholas and John Van Reenen (2001) "Real Options, Patents, Productivity and Market Value: Evidence from a Panel of British Firms", Department of Economics, University College London and Institute for Fiscal Studies
- Caballero, Ricardo and Adam Jaffe (1993) "Standing on the Shoulders of Giants: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth" in O. Blanchard and S. Fischer, eds, *NBER Macroeconomics Annual*, 1993
- Griliches, Zvi (1981) "Market Value, R&D, and Patents", *Economic Letters* vol 7, 183-187
- Griliches, Zvi (1984) "R&D, Patent, and Productivity", NBER Conference Proceedings, University of Chicago Press
- Griliches, Zvi, Ariel Pakes and Bronwyn H. Hall (1987) "The Value of Patents as Indicators of Inventive Activity", *Economic Policy and Technological Performance* Cambridge, England: Cambridge University Press, 97-124
- Hall, Bronwyn H, Jaffe, Adam and Trajtenberg (2000) "Market Value and Patent Citations: A first Look", NBER Working Paper 7741, May
- Hall, Bronwyn H., Adam B. Jaffe and Manuel Trajtenberg (2001) "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools", National Bureau of Economic Research Working Paper 8498, October
- Hall, Bronwyn H. (1998) "Innovation and Market Value" Prepared for the NIESR conference on Productivity and Competitiveness, London, February 5-6, 1998

- Harhoff, Dietmar, Francis Narin, F.M. Scherer and Katrin Vopel (1999) "Citation Frequency and the Value of Patented Innovations", *The Review of Economics and Statistics*, vol 81, 3, 511-515, August
- Hu, Alvert G. Z. and Adam B. Jaffe (2001) "Patent Citations and International Knowledge Flow: the Case of Korea and Taiwan", Department of Economics, National University of Singapore, Department of Economics, Brandeis University and NBER
- Jaffe, Adam and Manuel Trajtenberg (1999) "International Knowledge Flows: Evidence from Patent Citations", *Economics of Innovation and New Technology*, vol 8, 105-136
- Jaffe, Adam, Manuel Trajtenberg and Rebecca Henderson (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations", *The Quarterly Journal of Economics*, vol, 108, 3, 577-598
- Jaffe, Adam, Michael S. Fogarty and Bruce Banks (1997), "Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation", Working Paper 6044, National Bureau of Economic Research Working Paper .6044, May
- Jaffe, Adam, Manuel Trajenberg and M. Fogarty (2000) "Knowledge Spillovers and Patent Citations: Evidence from A Survey of Inventors", *American Economic Review*, 215-218, May
- Jaffe, Adam, Manuel Trajtenberg and Michael S. Fogarty (2000) "The Meaning of Patent Citations: Report on the NBER/CASE-Western Reserve Survey of Patentees", National Bureau of Economic Research Working Paper 7631, April
- Krugman, P. (1991) "Geography and Trade", Cambridge, M.I.T. Press
- Lanjouw, Jean O. and Mark Schankerman (1999), "The Quality of Ideas: Measuring Innovation with Multiple Indicators", National Bureau of Economic Research Working Paper No. 7345, September

Maurseth, Per Botolf (2001) "Lovely but Dangerous: The Impact of Patent Citations on Patent Duration", Norwegian Institute of International Affairs, Working Paper 614, March

Schmookler, J. (1966) "Invention and Economic Growth", Cambridge, MA: Harvard University Press

Stolpe, Michael (1998) "Inventor Mobility and Technology Diffusion as Evidenced in Patent Data the Case of Liquid Crystal Display Technology", Prepared for International Conference: Industrial Structure, Innovation Dynamics and Technology Policy Lisbon, 16-17, October

Shane, Hillary (1997) "Relation between Patent Citations and Tobin's Q in the Semiconductor Industry", *Review of Quantitative Finance and Accounting*, vol 9, 131-146

Scherer, F.M. (1965) "Firms Size, Market Structure, Opportunity and the Output of Patented Innovations", *American Economic Review*, vol 55 1097-1123

Trajtenberg, Manuel (1990) "A Penny for Your Quotes: Patent Citations and the Value of Innovations", *RAND Journal of Economics*, vol 21 172-187, Spring

Trajtenberg, Manuel (2000) "Is Canada Missing the Technology Boat? Evidence from Patent Data", Tel-Aviv University, National Bureau of Economic Research and Canadian Institute for Advanced Research Discussion Paper No. 9, January

Trajtenberg, Manuel (1989) "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners", *Journal of Political Economy*, vol 97 2, 445-479

Trajtenberg, Manuel (1990) "Economic Analysis of Product Innovation" Cambridge, Harvard University Press

Verspagen, B. (1997) "Measuring Intersectoral Technology Spillovers: Estimates from the European and US Patent Office Databases", *Economics Systems Research* vol 9, 47-65