FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

The Value of Knowledge Spillovers

Yi Deng
Southern Methodist University

June 2005

Working Paper 2005-14
http://www.frbsf.org/publications/economics/papers/2005/wp05-14k.pdf

The Value of Knowledge Spillovers*

Yi Deng†

June 2005

Abstract

This paper aims at quantifying the economic value of knowledge spillovers by exploring information contained in patent citations. We estimate a market valuation equation for semiconductor firms during the 1980s and 1990s, and find an average value in the amount of $0.6 to 1.2 million "R&D-equivalent" dollars for the knowledge spillovers as embodied in one patent citation. For an average semiconductor firm, such an estimate implies that the total value of knowledge spillovers the firm received during the sample period could be as high as half of its actual total R&D expenditures in the same period. This provides a direct measure of the economic values of the social returns or externalities of relevant technological innovations. We also find that the value of knowledge spillovers declines as the size of the firm's patent portfolio increases, and that self citations are more valuable than external citations, indicating a significant amount of tacit knowledge or know-how spillovers that occur within the firm.

*Financial support from Center for the Study of Innovation and Productivity at the Federal Reserve Bank of San Francisco is gratefully acknowledged.

†Department of Economics, Southern Methodist University, Dallas, TX 75275. Email: *ydeng@smu.edu*.

1. Introduction

Knowledge spillovers among different economic units are one of the most intriguing aspects of technological innovations and are of great importance for public policy making. Numerous studies have analyzed the patterns and the effects of such spillovers, at both microeconomic and macroeconomic levels (the endogenous growth theory by Romer (1986) and Grossman and Helpman (1991), for instance). However, we still know very little about how to quantify the economic value of such spillovers.[1] This study, by exploring firm-level data on patents, patent citations, R&D, and firm values, seeks to provide some answers.

The number of patent citations is used to measure the amount of knowledge flows. Patent citations, by identifying the previous relevant technologies on which the current patented technology builds, convey important information on knowledge spillovers that the current inventor has received from the earlier inventors. A number of authors have used patent citations to explore spillovers across geographical locations (Jaffe, Trajtenberg, and Henderson 1993), among firms in a research consortium (Ham, Linden, and Appleyard 1998), and spillovers from public research facilities to the whole economy (Jaffe and Trajtenberg 1996, Jaffe and Lerner 2001). This study proceeds along this approach and tries to quantify such spillovers in terms of monetary value, in an attempt to directly evaluate the social returns or externalities of the technological innovations as identified by previous studies.

The quantitative analysis of this paper is conducted in a Tobin's q framework. Following Hall, Jaffe, and Trajtenberg (2005) (hereinafter HJT 2005), a firm's knowledge assets are modeled as being accumulated in a continuously ongoing innovative process in which R&D expenditures reflect innovative input, patents record the successful innovations that can be appropriated by the firm, and citations *received* by the firm's patents (forward citations) measure the relative "importance" of the patents. The extension and fresh contribution of this study is to include the citations *made* by the firm's patents (backward citations) as a proxy of the knowledge flows the firm has received, which are considered an additional kind of innovative input to direct R&D spendings on the belief that more knowledge inflows increase the firm's knowledge stock and may boost the firm's R&D productivity.

[1] Trajtenberg (1990) estimates the social surplus from innovations in CT scanners technology based on a discrete-choice model, and finds high correlations between patent citations and the estimated social surplus.

Instead of a general analysis on all technology fields and industries aggregated, this study focuses on a very narrowly defined industry, the semiconductor industry (Standard Industrial Classifications or SIC code 3674). While analyzing an aggregate sample of different technology fields provides a broader picture, focusing on one industry allows more intensive and thorough examination of the technological competition and diffusion processes. The semiconductor industry is chosen because of the strategic importance of knowledge assets and the intensive R&D activities in this industry. Moreover, technological innovations in semiconductor industry have been the focus of several recent studies, including Megna and Klock (1993)'s work on R&D and patent stocks and Tobin's q, Ham, Linden, and Appleyard (1998)'s study on the knowledge spillovers in the research consortia in this industry, and Hall and Ziedonis (2001)'s extensive analysis of the shift in the patenting preferences of the semiconductor firms in early 1980s. This quantitative study on the economic value of knowledge spillovers in this industry based on patent citations data complements these previous studies and has a direct bearing on the existing literature.

The analysis examines the patenting behavior and market valuations of the universe of 120 semiconductor firms publicly traded in the U.S. during 1980s and 1990s, and find a significantly positive monetary value for knowledge spillovers. In particular, model estimation reveals an average value in the amount of \$0.6 to 1.2 million "R&D-equivalent" dollars for the knowledge spillovers as embodied in one patent citation, implying that the total value of knowledge spillovers a median-sized semiconductor firm received during the sample period could be as high as half of its actual total R&D expenditures in the same period. This provides a direct measure of the economic value of the social returns or externalities of relevant technological innovations. We also find that the value of backward citations declines when the size of the firm's patent portfolio increases, and that citations are more valuable for firms entering the semiconductor industry after 1982. In addition, estimation results also suggest that the technological competition in this industry can be quite intense so that the firms are unable to use patents to build and keep a leading position that is strong enough to effectively deter challengers. Finally, self citations are found to be substantially more valuable than external citations, indicating a significant amount of tacit knowledge or know-how spillovers that occur within the same firm.

The rest of the paper is organized as follows. Section 2 pictures the relationship between patent citations and knowledge spillovers, section 3 specifies the market value equation to be estimated, and section 4 describes the data. Section 5 presents the empirical results, and section 6 concludes.

2. Patent Citations and Knowledge Spillovers

Patent Citations as Indicators of Knowledge Spillovers

A patent grants it owner an exclusive right for the commercial use of the patented invention for a pre-determined period of time (20 years in the U.S.). Upon patent approval, a public document is created containing detailed information on various aspects of the invention and the inventor(s), including "references" or "citations." The citations serve an important legal function of delimiting the scope of the property right granted to the patent owner, since the patent only protects the exclusive use of the "novel and useful contribution over and above the previous state of knowledge, as represented by the citations" (Jaffe, Trajtenberg, and Henderson 1993, pp. 580). Thus, the cited patents represents a piece of previously existing knowledge upon which the citing patent builds, and over which the citing patent *cannot* have a claim. The patent applicant has a legal duty to disclose any knowledge of prior art, although the patent examiner will make the final decisions on which previous patents to be included in citations.[2]

The fact that patent citations reveal the "prior art" the inventor has learned makes them potential measures of the knowledge spillovers from the past inventions to the current invention. Undoubtedly, there is substantial noise in using patent citations to measure knowledge spillovers. To assess the validity of this analysis, let us first examine the relationship between patent citations and knowledge spillovers more carefully. There are three possibilities: spillovers accompanied by citations, citations that occur where there is no spillover, and spillovers that occur without generating a citation (Jaffe, Trajtenberg, and Henderson 1993). The validity of this empirical analysis relies on the first one. So the key question here is whether and to what

[2] As noted by Jaffe, Trajtenberg, and Henderson (1993), one should be careful in making analogy between patent citations and academic article citations: the price of making an academic citation is almost zero or even negative if a long list of references may make the research papers seem more solid and thorough. However, under the patent system, the more citations a patent applicant makes, the less novelty or significance he is able to claim over his invention. Therefore, the patent applicant may have an incentive to *under*-cite rather than *over*-cite the precedents, and the patent examiner will use his expertise to identify the ones neglected by the applicant. On the other hand, even if the inventor did cite some irrelevant patents in the application (which is rare given his incentives), the examiner should exclude them in the final grant.

extent the other two possibilities may affect the evaluation of spillovers.

A recent survey study of inventors provides some direct evidence. Jaffe, Trajtenberg, and Fogarty (2000) interviewed approximately 160 patent owners with questions about their inventions, the relationship of their patents to the patents they cited, as well as the relationship to other patents that were technologically similar to the cited patents but not cited. The study concludes that about half of the citations correspond to some knowledge flows from the cited patents to the citing patents, and the other half does not seem to correspond to any kind of knowledge flow between them. This confirms that citations do contain important information about knowledge spillovers ("spillovers accompanied by citations"), but with a substantial amount of noise ("citations that occur where there is no spillover"). This implies that the estimated value of patent citations will reflect a lower bound of the economic value of knwledge spillovers.[3][4]

Meanwhile, there are an enormous amount of spillovers not reflected in patent citations, since only a fraction of research output is patented ("spillovers that occur without generating a citation"). For instance, results from basic research are seldom patented, although they may generate huge amount of spillovers. Thus, our estimation results are more relevant to the applied research than basic research, and imply a lower bound of the total value of spillovers *received* by the whole industry. On the other hand, since the R&D projects in semiconductor industry are more oriented toward applied rather than basic science and technology, especially the ones funded and owned by private firms, our estimates of spillovers *within* the semiconductor industry will suffer less from bias of this kind.

Knowledge Spillovers Within and Beyond the Firm and the Industry

In the following empirical analysis we classify the patent citations into three groups: citations occur within the same firm (self citations), external citations to other semiconductor patents, and citations to non-semiconductor patents. Such distinction is made because the economic value of each kind of citations may be different, for two reasons:

[3] In principle, such an "proxy variable" problem could be solved by conducting an instrumental variables (IV) estimation and general point estimates of the value could be obtained. However, we are lack of reliable instrument variables for backward citations (ideal instrumental variables should be highly correlated with knowledge spillovers but not with citations measurement errors). Thus we do not pursue this approach in the current analysis.

[4] Royalty payments from unilateral licensing agreements apparently provide another possible estimate of the lower bound of knowledge spillovers between firms. But such data are not readily available. Moreover, it is the cross-licensing agreements that are prevalent in semiconductor industry, in which no net royalty payment is made.

First, the content and thus the economic value of knowledge transfer as represented by each type of citations may be different. When applying for a patent, the inventor has some discretion over how to codify and disclose the new knowledge (Arora and Fosfuri 1998). He may choose not to disclose every piece of the new knowledge and keep part of it "tacit," in order to discourage potential followers in the same or similar technological areas (or imitators when the patent protection is not perfect). Therefore a self citation reflects an internal transfer of both the codified and the tacit knowledge ("know-how"), while an external citation to another firm may only indicate a transfer of codified knowledge but not know-how, and the difference between the estimated value of internal and external citations may reveal the value of know-how transfers.[5]

Secondly, in the process of sequential innovation in the same narrow technology field, successive inventors compete away each other's excess returns (Scotchmer 1991). In that sense, a self citation would imply that the firm is now gaining a more competitive position in that field, while an external citation to another semiconductor patent may suggest that the citing firm is entering a technological competition where the cited firm might have already built a strong competitive position. On the other hand, the knowledge flows from a non-semiconductor patent (as embodied in an external citation to that patent) may have much weaker implications on the technological competition, because the cited patent is in a different technology field. Thus making distinctions between these two kinds of citations may shed light on the intensity of the technological competitions in the semiconductor industry.

3. Model Specification

Consider the following market valuation equation from Griliches (1981):

$$V_{it} = q_t(A_{it} + bK_{it})^{\sigma} \tag{3.1}$$

where V_{it} denotes firm i's stock-market value in year t, A_{it} the book value of its physical assets, and K_{it} the knowledge assets. q_t represents the shadow value of firms' assets, and the coefficient

[5] This distinction is somehow obscured by the knowledge transfers between different firms in a cross-licensing agreement or between collaborators within the same research consortium, which are prevalent in semiconductor industry. It is likely that in such occasions not only the codified knowledge but also some tacit knowledge is shared. Therefore, some observed patent citations between different semiconductor firms may also include a know-how transfer. However, there has not been a reliable and complete record of all the semiconductor firms involved in such cross-licensing agreements or research consortia.

b measures the shadow value of knowledge assets relative to physical assets. σ measures the scale effects in the value function and is often assumed to be one.

Taking logarithm of equation (3.1) yields

$$\log V_{it} = \log q_t + \sigma \log A_{it} + \sigma \log(1 + b\frac{K_{it}}{A_{it}}) \tag{3.2}$$

When the value function exhibits constant returns to scale (which holds approximately in the cross section), we have the following estimation equation:

$$\log Q_{it} = \log(\frac{V_{it}}{A_{it}}) = \log q_t + \log(1 + b\frac{K_{it}}{A_{it}}) + \varepsilon_{it} \tag{3.3}$$

where Q_{it} denotes Tobin's q. ε_{it} represents the prediction errors.

There is little guidance in theory on the specification of knowledge assets K_{it}. Accumulated knowledge spillovers into the firm directly increase the firm's knowledge base and boost the firm's R&D productivity, thus the accumulated backward citations, as a proxy of the spillovers, should be included. On the other hand, the literature has found that the accumulated R&D spendings are quite effective in predicting the market value of the firms. This is not surprising, as the accumulated R&D spendings measure the past efforts the firm has made in inventive activities, and even if some of the R&D projects turn out to be "dry holes," the spendings on those projects still increase the firm's knowledge assets through building firm's know-how. Therefore the accumulated R&D expenditures should also be included.

In addition, there is a high degree of heterogeneity in the R&D productivity across different firms. This heterogeneity also should be taken into account, because the market will use information on a firm's R&D productivity to evaluate its knowledge assets. A natural choice of R&D productivity measure is the number of patents owned by the firm, as patent counts indicates the "success" of R&D projects and thus the patent/R&D ratio measures the R&D productivity, similar to an output/input ratio (Scherer 1965, Griliches 1984, among others). However, the quality and value of different patents varies a lot, and the raw patent counts simply ignores this heterogeneity. HJT (2005) suggest using the number of forward citations (citations *received* by the patent) to weigh the patent counts and refine this measure, as a more frequently cited patent is technologically more important than other patents and potentially more valuable.

Thus, the market is assumed to use the following value function to evaluate the firm's knowledge assets

$$K_{it} = f(R\&D_{it}, BCIT_{it}, \omega_{it}) \tag{3.4}$$

where $R\&D_{it}$ denotes the accumulated R&D spendings, $BCIT_{it}$ the accumulated backward citations the firm has made as a proxy of the knowledge inflows received by the firm, and ω_{it} the accumulated idiosyncractic productivity shocks in the firm's inventive activities.

Taking first-order Taylor expansion of equation (3.4) yields

$$K_{it} = f_1 * R\&D_{it} + f_2 * BCIT_{it} + f_3 * \omega_{it} \tag{3.5}$$

As there is no directly observable measure of the idiosyncractic productivity shocks ω_{it}, we adopt HJT (2005)'s specification and proxy it by the patent/R&D ratio, weighed by the average number of forward citations the firm's patents receive over their entire lives (30 years after applications). This could be viewed as an approximate measure of the output-input ratio in the firm's R&D production.

Thus equations (3.3) and (3.5) imply the following basic estimation equation

$$\log Q_{it} = \log q_t + \log(1 + b_1 \frac{R\&D_{it}}{A_{it}} + b_2 \frac{BCIT_{it}}{A_{it}} + b_3 \frac{PAT_{it}}{R\&D_{it}} + b_4 \frac{FCIT_{it}}{PAT_{it}}) + \varepsilon_{it} \tag{3.6}$$

where PAT_{it} and $FCIT_{it}$ are firm i's patent stock and forward citations stock in year t. Here b_2 represents the value of knowledge flows brought by an additional backward citation, and b_2/b_1 is a direct measure of the monetary value of knowledge spillovers in terms of "R&D equivalent dollar." A full model estimation will further categorize the backward citations stock into stocks of self citations, external citations to other semiconductor patents, and external citations to non-semiconductor patents, as discussed in Section 2.

4. Data

Empirical estimation is based on the universe of 120 semiconductor firms publicly traded in the U.S. during 1979 and 1998. The final estimation sample only include firms whose primary business is in SIC 3674 (semiconductors and related devices). Conglomerates whose principle products are not semiconductors such as IBM, AT&T are excluded — although these firms are heavy users and important owners of semiconductor patents, we are unable to observe the R&D resources primarily devoted to semiconductor-related R&D projects by them, nor the market valuation of their semiconductor division.

The sample is constructed by combining information from two data sources. For market valuation of firms we use the popular Compustat database. As market value and book value of the firms are readily available in Compustat, the calculation of Tobin's Q is quite easy and straightforward. R&D expenditures, based on which R&D stocks are constructed, are also obtained from Compustat.

For patent and patent citations, we use the U.S. Patent Citations database, recently constructed by Hall, Jaffe, and Trajtenberg (2001). The database keeps a complete record of citations made by each U.S. patent upon approval since 1975, as well as other patent characteristics such as application date, approval date, and detailed classification code describing the technological classifications of the patent. For the purpose of this study, we first identify all the patents owned by each of the 120 semiconductor firms,[6] and for each patent we count the backward and forward citations for each year; then aggregate them on the firm level to construct the patent stocks as well as backward and forward citations stocks.

Dealing with Truncation[7]

There are two kinds of data truncation problems that one may encounter in constructing the patent and patent citations stocks. The first regards the patent counts and the backward citations counts. There is substantial time delay in granting of the patents: the average and median length of patent application review in the U.S. are approximately two years. Therefore, for the last two years of the Patent Citations database (which ends in 1999), one can only observe a fraction of the total patents that will be finally granted, as a lot of them were still being examined by the end of 1999 and were therefore not included in the database. In the estimation we solve this problem by focusing on a sample period that ends in 1995 — a preliminary look of the U.S. semiconductor patents indicates that, over the past three decades, 95% of the grant decisions on these patents were made within three years since the initial applications, and within four years more than 98% of the decisions were reached. Our selection of sample period guarantees that, even for the last year of the sample, at least 98% of the granted patents and

[6] Because the patent assignees obtain patents under a variety of names, and the US PTO does not keep a unique identifier for each patenting organization, we have performed an extensive name-matching exercise to identify the patent assignees in the citation daba base and link them to the firm names in the Compustat. The subsidiary relationship is identified according to the Directory of Corporate Affiliations, and keep track of major mergers and acquisitions events according to CRSP database.

[7] We follow the pioneering work of Hall, Jaffe, and Trajtenberg (2005) in dealing with the truncation problem on patent citations data as well as in constructing the patent citations stocks. For details please refer to the appendix.

backward citation counts is included, and thus keeps this truncation problem to a minimal degree.[8]

The second truncation problem concerns the forward citation counts and is due to the time lags in observing the forward citations. Such lags can be very long, as it is not unusual for a patent to be cited 10 or even 20 years after its initial application. Since the Patent Citations database ends in 1999, it only has a truncated record of the forward citations: for instance, for a patent belonging to cohort 1985 (i.e., application submitted in 1985), we are only able to observe 14 years of its forward citations history, and for a 1995 patent, only 4 years.[9] To address this problem, we follow Hall, Jaffe and Trajtenberg (2004) and estimate a citation-lag model. Based on the model estimates and conditional on the number of forward citations as observed in the data, we then project the number of forward citations received by each patent for the years not observed in the database, up till 30 years after their initial applications. The details of the citation-lag model estimation and projection are laid out in the appendix.

A by-product from estimating the structural citation-lag model is that we can parse out an important time effect on the overall changes in citation practice since mid-1970s. Preliminary data analysis indicates that the average number of citations made by each patent in the sample increased substantially during the sample period, from 3.9 in 1975 to 5.1 in 1985 and 10.1 in 1995. This increasing trend may not necessarily reflects a similar increase in the *substance* of

[8] In fact, at the beginning of the sample period one may also encounter some truncation problem regarding the patent counts and the backward citations counts: as the U.S. PTO did not begin to keep records of patent ownership until 1969, all the patents that can be identified as owned by the 120 semiconductor firms in the sample were granted after 1969 but none before. That is, for those firms who existed and possessed patents before 1969, their patent stocks are under-estimated in the sample. However, the total number of semiconductor patents granted before 1969 is quite small (for instance, only 372 semiconductor patents were granted in 1967 and 376 patents in 1968), and under an annual depreciation rate of 15%, the mis-measurement of the accumulated patent stock and backward citations stocks for years after 1979 is tiny.

[9] What makes it worse is that, for relatively younger patents, most of their citing patents had not yet been granted by 1999. For instance, for a 1996 patent, we only observe a fraction of total forward citations from cohort 1998 patents, as more than half of cohort 1998 had not been granted by 1999 and are thus excluded from the database. So for this patent we only have a reliable citations record of the first year after its approval, at most. This is another reason why we restrict the sample period to end in 1995, as 95% of cohort 1996 and 80% of cohort 1997 had been granted by 1999. Therefore even for the latest cited cohort in our sample (1995), we are still able to observe at least a couple of years of reliable forward citations records, based on which we can then project the life-long forward citation counts, as explained later.

knowledge flows an average inventor receives, bur rather partly due to some technical reasons. For instance, with the development of machine-readable patent databases and more accessible patent-searching tools over this period, the patent attorneys and examiners are better equipped to identify the relevant previous patents in making citations. If so, this "citation inflation" would imply that a typical backward citation indicates less amount of actual knowledge spillovers in 1990s than in 1970s. We thus construct two samples, one "deflated" sample making adjustments for this "citation inflation" (on both backward and forward citations) and one "undeflated" sample without making such adjustments.

Construction of R&D, Patent and Patent Citations Stocks

The construction of R&D stock is fairly straightforward, as it is simply accumulated past R&D expenditures. Therefore,

$$R\&D_{it} = \sum_{j=0}^{t-T_0} \delta^j * r\&d_{i,t-j} \tag{4.1}$$

where $r\&d_{i,t-j}$ is the R&D spendings by firm i in the year of $t-j$. δ is an annual depreciation rate assumed to be a constant 15%, as in much of the literature. T_0 is the beginning of the database. The patent stock is defined in the same fashion.

Knowledge that the firm acquires in the past also depreciates over time. We depreciate the number of patent citations according to the age of the cited patents (throughout the paper the age of a patent is defined as the time elapsed since the patent *application*). For instance, if a firm cites a 1975 patent in one of its 1980 patents, then the knowledge that the firm learned in 1980 from the 1975 patent was already 5 years old and needs to be discounted (subject to the same 15% depreciation rate). As time goes by, the value of that piece of knowledge continues to depreciate, and in 1990, it will be 15 years old and worth only 8.7% as a new citation made to a 1990 patent. The step aggregates such accumulated backward citations over the firm's patent portfolio each year, to obtain the firm's backward citations stock.

The forward citations stock measures the relative importance or value of the firm's patent portfolio. For each year, we aggregate over the entire patent portfolio the number of forward citations received by each patent during its entire life (30 years since application), and discount them according to the age of the patents. For instance, suppose the firm has one 1980 patent and one 1985 patent which are projected to receive 10 and 8 citations during their entire lives, respectively. Then in computing the forward citations stock for the firm in 1990, we would discount the forward citations received by first patent by 10 years as $10 \times 0.85^{10} = 1.97$ and those received by the second patent by 5 years as $8 \times 0.85^5 = 3.55$, so the entire forward citations stock is $1.97 + 3.55 = 5.52$ in 1990. In other words, we do not distinguish as to when the forward citations arrive, but rather discount the sum of them according to vintage of the cited patent.

First look at the Sample

Market valuation of semiconductor firms can be quite volatile.[10] To reduce the idiosyncractic shocks especially from young start-up firms, we eliminate firms with less than three years of complete observations in the Compustat from the sample. We also delete several observations in which Tobin's q seems too high (greater than 10). This generates a sample of 64 firms (possessing a total of 26,143 patents during the sample period) in an unbalanced panel, or 636 firm-year observations.[11] Table 1 presents some summary statistics of the estimation sample. The market value and the book value of the firms are extremely skewed to the right, with means several times larger than medians. The skewness is even more pronounced for all the determinants of knowledge stocks (R&D, patents, backward and forward citations), with the means usually ten times larger than medians or more. On the other hand, variables such as Tobin's Q, R&D stock/total assets, backward citations/total assets, patents/R&D and forward citations/patents are much more symmetrically distributed, with means usually only twice as large as the medians or even less. Finally, about 14 percent of all the firm-year observations have a zero patent stock. Therefore we also construct another "patenting" sample of 545 firm-year observations whose patent stocks are positive.

Table 1 also indicates that the mean and median of the projected lifetime forward citations stock are several times larger than those of backward citations stock. This is surprising as in the long run these two measures should be comparable. The discrepancy between them is closely related to the rapid growth of the number of patents in semiconductor industry since 1970s (both because of the rapid expansion of this industry and a shift in the patenting prefer-

[10] For instance, during the stock-market bubbles in late 1990s, the market values of semiconductor firms were substantially blown up, in many cases by several times. This is another reason why we choose to let the sample period end in 1995 and delete those young start-up firms, in order to minimize the distortions in market valuation of the firms.

[11] The eliminated firms are either small start-ups short of three years of public-trading history or firms without three years of complete trading data, or foreign firms, and they only possess a total of 1,789 patents during the sample period.

ences of the semiconductor firms starting in early 1980s (Hall and Ziedonis 2001)): even if each patent makes the same number of backward citations over time, the average number of forward citations received by each earlier patent may be much larger than the average number of backward citations, simply because now there are more later patents citing earlier patents. This is another kind of "citation inflation" indicating that in an industry where the total number of patents change substantially over time, the forward citations count is a more "noisy" measure of technological or economic "importance" than in other industries.[12] The distortion on backward citations count, on the other hand, is quite small.

The top panel of Table 2 shows that R&D, patents, backward and forward citation stocks are highly correlated with each other, with the correlation between R&D and patent stocks being 0.83, and that between backward and forward citations even higher. This is not surprising, since all of them are different measures of knowledge stock. However, the correlations between different regressors of the estimation equation such as R&D/total assets, backward citations/total assets, patents/R&D, and forward citations/patents, are much lower (less than 0.5), indicating that each of the regressors possess independent information content and the colinearity problem is not severe.

5. Estimation Results

The market value equation (3.6) and its variants are estimated using maximum likelihood estimator. Year dummies are included to allow Tobin's q to vary over time.

First take of model estimation

Table 3 displays the estimation results of equation (3.6) based on two samples: the top panel uses all 636 firm-year observations and the bottom panel focuses on the 545 firm-year observations with positive patent stock. For each sample we start by regressing the market value on R&D stock, and then gradually add other regressors to the equation, one at a time. This procedure facilitates the examination of the significance and the marginal contribution

[12] Adding a time trend in the citation-lag model estimation, either a linear trend or some other kind of filtered trend of the growth rate of patent number over time may help solve this problem. In this paper we do not pursue this possibility, as the focus here is on knowledge spillovers proxied by the backward citations, which are much less affected by this kind of distortion. Moreover, although this inflation on the number of forward citations tend to decrease the average value of each forward citation, the distortion is much less on the total value of all the forward citations that a patent receives.

of each regressor. As discussed in Section 4, we make distinctions on whether the "citation inflation" is adjusted when constructing the citations stock, and run separate estimations for each case.

All the coefficient estimates in Table 3 are positive and significant at 5 percent level, suggesting that all regressors have significant impact on market value. Moreover, the likelihood ratio tests indicate that all regressors adds information on top of others, and thereby have a significant contribution to the overall fit of the estimated model.

Most of the coefficient estimates are very similar in both panels and when backward citations are not included in the equations, are quite close to those previously estimated for a broader set of industries, for instance HJT (2005)'s estimates on computer sector (where backward citations are not included as well). In particular, the coefficient estimate of R&D/assets is around 0.26, close to 0.32 in HJT (2005); and of patents/R&D is around 0.13, compared with 0.06 in HJT (2005). The coefficient estimate of forward citations/patents, on the other hand, is about 0.003 to 0.004 in the table, much smaller than the estimate of 0.028 in HJT (2005) at the first look. However, it should be noted the average size of forward citations stock/patents in sample is much larger — a mean of 68 for patenting firms in Table 1 versus a mean of 8 in their patenting sample, or a median of 33 versus 6.3. As discussed in Section 4, this probably comes from the citation inflation associated with the rapid increase in the number of patents in semiconductor industry over the past two decades. This distortion in forward citations counts leads to a decrease in the estimated value of each forward citation, but the total value of forward citations for each patent, when summed up over all the forward citations it receives, is close to their estimates.

Column (4) of the table indicates that the coefficient estimates of R&D/assets and patents/R&D decline when backward citations/assets is added to the equation: the R&D/assets coefficient falls from 0.26 to around 0.17, and the patents/R&D coefficient falls from 0.13 to 0.10. The coefficient estimate of backward citations/assets is significantly positive and is around 0.08 when citation inflation is not adjusted or 0.10 when citation inflation is adjusted in the top panel, and even higher (0.12) in the "patenting sample" estimation as shown in the bottom panel.

Next we examine the quantitative impact of the regressors on market value using these coefficient estimates. Consider the following semi-elasticity:

$$\frac{\partial \log Q}{\partial (R\&D\ /\ A)} = \widehat{b}_1(1 + \widehat{b}_1 \frac{R\&D_{it}}{A_{it}} + \widehat{b}_2 \frac{BCIT_{it}}{A_{it}} + \widehat{b}_3 \frac{PAT_{it}}{R\&D_{it}} + \widehat{b}_4 \frac{FCIT_{it}}{PAT_{it}})^{-1} \tag{5.1}$$

which provides a rough estimate of the elasticity of Tobin's q with respect to an increase in

$R\&D\ /A$ ratio (HJT 2005). We evaluate this elasticity around the mean and median value of the regressors as in Table 1, based on the estimated coefficients in column (4) in the top panel (all firms, with patent citations not deflated) and column (6) in the bottom panel (patenting firms, with citations deflated) of Table 3. Elasticities with respect to changes in other regressors are also calculated and displayed in Table 4. Calculations based on the other two sets of coefficient estimates (column (6) in the top panel and (4) in the bottom panel of Table 3) are quantitatively similar and thus not reported.

As shown in Table 4, a one-percentage point increase in R&D/assets ratio leads to a 0.1% appreciation in the firm value. An increase in the firm's R&D productivity, as measured by one extra patent per million dollar R&D spendings, boost the firm value by 6% to 7%, about three times as high as the elasticity of 2% for all manufacturing sector as estimated by HJT (2005). This is consistent with the strategic importance of patents in semiconductor industry (Hall and Ziedonis 2001). A rise in the average quality of the firm's patent portfolio also raises the firm's market value — if every patent receives one more forward citation over their entire lives (30 years since applications), the firm's value will rise by about 0.3%.

Of particular interest to us is the impact of backward citations on firm value as it proxies the value of knowledge spillovers. As displayed in Table 4, one extra backward citation per million dollar of physical asset makes the firm about 5% more valuable, and the amount of appreciation is even larger (7.5% to 9%) when the number of citations is deflated.

Another way to quantify the value of spillovers is to calculate how much R&D spendings has to be increased in exchange for one less backward citation, keeping the firm value unchanged $(\widehat{b}_2/\widehat{b}_1)$. Estimates in Table 3 indicates that this figure ranges between \$0.6 million and \$0.7 million (in 1998 value). This translates into a total value of about \$12 million for a firm with a median size of backward citations stock (about 20 as in Table 1), which is about half of the accumulated R&D stock for a median firm (\$26 million).

Controlling for Firm Characteristics

In Table 5 we control for several firm characteristics that may also affect the firm value and the value of knowledge spillovers. At first the logarithm of net sales of the firm is included to examine the impact of firm size (columns (2) and (7)). We then also introduce a dummy on whether the firm entered the industry after 1982 ("post-82 entrant"). As Hall and Ziedonis (2001) point out, semiconductor firms that entered the industry after 1982 have a significantly higher tendency to seek for patent protection for their inventions than firms entering before 1982, because of the more "pro-patent" legal environment (the creation of Court of Appeal for Federal Circuit in 1982 and other legal changes in early 1980s such as the Semiconductor Chip Protection Act in 1984). We also include another dummy for Texas Instruments Inc. ("TI"), for its well-known strategy of aggressively pursuing patent protection as well as its large size of patent portfolio (the firm alone owns about 30% of all the patents in the sample).

Columns (2) and (7) of the table indicate a slightly positive premium for larger semiconductor firms, although the premium is not statistically significant and diminishes when "TI" and "post-82 entrant" are added (columns (3) and (8)). Texas Instruments has a significantly negative premium on market value which lowers its Tobin's q by about 20% to 25%. On the other hand, firms entering the semiconductor industry after 1982 have a significantly positive premium, in the amount of about 60% of the firm value.

In Table 5 we also interact the log sales and "post-82 entrant" with backward citations/assets and examine how differently the knowledge flows are valued in different types of firms. Columns (4) and (9) show that knowledge flows are evaluated quite differently in regard to the timing of firm's entrance into the semiconductor industry. For older firms entering the industry before 1982, each backward citation is worth about \$0.6 million "R&D equivalent dollars," whereas for those entering the industry after 1982, the value is about one and half times larger, at about \$1.4 million. In other words, younger firms are not only more prone to patenting (Hall and Ziedonis 2001), they also appear to benefit more from the knowledge spillovers.

Columns (5) and (10) show that the larger the firm size, the less valuable the backward citations are. For instance, for a median-sized firm (with net sales around \$110 million each year), each backward citation is worth about \$0.5 million or \$1 million, depending on whether citations stock is deflated, and for a firm whose net sales are at the top 25 percentile (\$328 million annually), each backward citation is only worth \$0.06 million (column (10)) or less.

Because the larger firms in semiconductor industry usually hold more patents, to single out the true "firm-size" effect apart from "patent-portfolio size" effect, we also include in the estimation equations the size of the firm's patent portfolio (defined as the raw count of patents that the firm had ever acquired), and present the estimation results in Table 6. It is found that firms with larger patent portfolio value backward citations less, as the coefficient estimate of BCIT/assets interacted with patent portfolio size is significantly negative (columns (2) to

(5) and (7) to (10)). Moreover, the coefficient on net sales becomes insignificant, even when net sales is interacted with the BCIT/assets (columns (4) and (9)). This indicates that the "firm-size" effect on the value of knowledge spillovers in Table 5 is indeed spurious and simply reflects a negative "patent-portfolio size" effect and the fact that larger firms usually hold more patents. The positive premium on the value of spillovers for post-82 entrants, however, remains significant (columns (4), (5), (9) and (10)). Finally, when the patent portfolio size is included in the equation, the estimated coefficient of the dummy for Texas Instruments Inc. becomes positive because of the huge number of patents the firm possesses.

Table 7 presents a direct look into how the estimated value of backward citations decline as the size of patent portfolio increases. For firms with patent portfolio size at the lowest 25 percentile (holding 8 patents), each backward citation is worth approximately \$1.5 million or \$2.9 million depending on when the firm entered the industry; for firms with median-sized patent portfolio (28 patents), the average value of backward citations is \$0.4 million or \$1.7 million; and for firms with a patent portfolio size at the top 25 percentile (holding 95 patents), the average value is much less, lower than \$0.6 million for post-82 entrants and even less for firms entered the industry before 1982. This suggests that the knowledge spillovers are more valuable for younger firms with few patents, and for older firms with a large patent portfolio, the unit value of knowledge inflows added on top of their already abundant knowledge base is relatively smaller.

Spillovers within and beyond the firm and industry

Finally we make distinction between backward citations within and beyond the firm and industry, as they may differ in the amount of knowledge flows they carry (whether including tacit knowledge or not), and on the implications for technological competitions they may have. In particular, we distinguish self citations (citations to other patents the same firm owns) from external citations, and further divide the external citations into two groups: citations to non-semiconductor patents and those to semiconductor patents. Accordingly, two new variables are constructed: the non-semiconductor backward citations stock/assets (NSCBCIT/assets), and SelfBCIT/BCIT representing the share of self citations stocks in the total backward citations stocks of the firm. As the backward citations to non-semiconductor patents enter both BCIT/assets and NSCBCIT/assets, the estimated coefficient of NSCBCIT/assets is indeed a premium over backward citations to semiconductor patents. Similarly, a self citation enters both BCIT/assets and SelfBCIT/BCIT, so the estimated coefficient of SelfBCIT/BCIT measures the premium of self citations over backward citations to external semiconductor patents. Table 8 presents the estimation results based on the "patenting sample", when all citations are deflated (estimates based on other three cases are similar and thus not reported).

First let us focus on the differences between external citations made to patents within the same semiconductor industry and those made to patents from other industries. Columns (1) and (2) of the table show that, among the external citations that a firm makes, citations to non-semiconductor patents tend to have a lower average value than citations to semiconductor patents, but the differences are not statistically significant. As discussed in Section 2, external citations to semiconductor patents may imply that the citing firm is entering a technological competition with the cited firms who might have already built a strong leading position in that area, whereas such implications are much less relevant to external citations to non-semiconductor patents as the cited patents are less likely to be competing with the citing patent, and the difference between the estimated value of these two kinds of citations may shed light on the intensity of technological competition in this industry. If the leading position of cited semiconductor patent is very strong, then we should expect a significantly positive premium on the value of non-semiconductor citations over that of semiconductor citations.[13] Thus, the insignificant estimate of this difference suggest that the disadvantage of being late in the technological competition in semiconductor industry is not significant. In other words, the incumbent firms (which possess earlier patents in this area) are not able to build and keep a position that is strong enough to effectively deter other firms from entering the competition. This is consistent with the rapid technology pace in this industry and more importantly, the fact that technological innovations in this industry are a "cumulative" process (Levin, Klevorick, Nelson, and Winter 1987, Scotchmer 1991) in which innovations are built successively on previous inventions and therefore often require access to hundreds of patents owned by a diverse set of entities (Hall and Ziedonis 2001).

[13] The value of these two kinds of citations may differ for another reason, that the knowledge embodied in a non-semiconductor backward citation maybe less technologically relevant to the firm than a citation to semiconductor patent, and thus is less valuable. However, we check the technological classification code of those cited non-semiconductor patents, and find that the majority of them are in quite relevant technology fields, such as "electrical computers and digital data processing systems" (codes 710, 711, and 712, which include processors and memory) and "static information storage and retrieval" (code 365), etc. Therefore, the differences in the technological relevance of the knowledge flows as embodied in these two kinds of backward citations should not be very large.

Cumulative innovations, rapid change, and multiple owners of overlapping technology rights make it very difficult to build and keep a leading position that is strong enough to effectively deter challengers.

Columns (3) and (5) of Table 8 reveal how the values of these two kinds of external citations vary with the size of the firm's patent portfolio. The value of non-semiconductor citations do not vary a lot as the size of the firm's patent portfolio increases. However, the value of semiconductor citations decline as the firm holds more patents. This seems to suggest that, when the knowledge spillovers occur within the same technological field, the value of such spillovers decline with the size of the receiving firm's knowledge base, as firms with large patent portfolio may have already accumulated a lot of similar knowledge in that area so each additional unit of knowledge inflows becomes less valuable; on the other hand, firms with different patent portfolio sizes may be equally unfamiliar with knowledge outside their expertise, and thus the value of citations to such patents does not vary much.

Next we explore the implications on the value of tacit knowledge as implied by the coefficient estimates on self citations. The estimated coefficient of self-citations stock/total citations stock (SelfBCIT/BCIT) is significantly positive in columns (1) to (3). Since self citations are also included in total backward citations, this coefficient estimate represents the premium of self citations over external semiconductor citations. In particular, column (2) indicates that, for a median semiconductor firm, a 10-percentage point rise in the share of self citations increases the firm value by about 6%. And the premium over external non-semiconductor citations may be even higher.

The estimated positive premium on self citations is consistent with the conjecture in Section 2 that self citations are more valuable to the firm because of the additional tacit knowledge or know-how transfer that took place within the firm, as well as a value in the strengthening of the firm's position in the technological race. As we have learned that the latter seems to be relatively small in this industry, the bulk of the positive premium on self citations would be the value added brought by the internalized know-how spillovers. For a median firm, this translates into a monetary value of about $0.4 million for firms entering the industry after 1982, or $0.28 million for firms entering before 1982. Moreover, columns (4) and (5) show that such premiums increase as the size of the firm's patent portfolio increases, suggesting a higher load of tacit knowledge for firms with more patents. And the value of such internalized spillovers are significantly higher for post-82 entrants than older firms.

6. Concluding Remarks

This paper aims at quantifying the economic value of knowledge spillovers by exploring information contained in patent citations. We estimate Tobin's q equations on various determinants of semiconductor firms' knowledge assets, and find an average value in the amount of \$0.6 to 1.2 million "R&D-equivalent" dollars for the knowledge flows as embodied in one patent citation. For an median semiconductor firm, this implies that the total value of knowledge spillovers it had received during the sample period could be as high as half of its actual total R&D investment during the same period.

We also find that the value of backward citations decline when the size of the firm's patent portfolio increases, and that citations are more valuable for firms entering the semiconductor industry after 1982. In addition, estimation results suggest that the technological competition in this industry can be quite intense so that the firms are unable to use patents to build and keep a leading position that is strong enough to effectively deter challengers. Finally, self citations are found to be substantially more valuable than external citations, indicating a significant amount of tacit knowledge or know-how spillovers that occur within the same firm.

References

[1] Arora, Ashish and Andrea Fosfuri, 1998. "Licensing in the Chemical Industry," working paper, Heinz School of Public Policy and Management, Carnegie Mellon University, 1998-24.

[2] Griliches, Zvi, 1981. "Market Value, R&D, and Patents," *Economic Letters*, Vol. 7, pp. 183-187.

[3] —— (ed.), 1984. *R&D, Patents and Productivity.* University of Chicago Press, Chicago.

[4] ——, 1990, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, Vol. XXVIII, 1661-1707.

[5] —— and J. A. Hausman, 1986. "Errors in Variables in Panel Data," *Journal of Econometrics*, vol. 31, pp. 93-118.

[6] Grossman, Gene and Elhanan Helpman, 1991. "Quality Ladders and Product Cycles," *Quarterly Journal of Economics*, vol. 106, pp. 557-586.

[7] Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, 2001. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," NBER working paper No. 8498. Reprinted as Chapter 13 in Jaffe and Trajtenberg (200

[8] ——, ——, and Manuel Trajtenberg, 2005. "Market Value and Patent Citations," *Rand Journal of Economics*, vol. 36, No. 1.

[9] —— and Rosemarie Ham Ziedonis, 2001. "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995," *Rand Journal of Economics*, vol. 32, No. 1, pp. 101-128.

[10] Ham, Rose Marie, Greg Linden, and Melissa M. Appleyard, 1998. "The Evolving Role of Semiconductor Consortia in the United States and Japan," *California Management Review* vol. 41, pp. 137-163.

[11] Harhoff, Dietmar, Francis Narin, F.M. Scherer and Katrin Vopel (1999), "Citation Frequency and the Value of Patented Inventions," *The Review of Economics and Statistics*, 81, 3, 511-515.

[12] Jaffe, Adam and Joshua Lerner, 2001. "Reinventing Public R&D: Patent Policy and the Commercialization of National Laboratory Technologies," *Rand Journal of Economics*, vol. 32, No. 1, pp. 167-199.

[13] —— and Manuel Trajtenberg, 1996. "Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundaries," *Proceedings of National Academy of Science*, vol. 93, pp. 12671-12677.

[14] —— and ——, 2002. "Patents, Citations and Innovations: A Window on the Knowledge Economy." Cambridge: MIT Press.

[15] ——, ——, and Michael Fogarty, 2000. "The Meaning of Patent Citations: Report of the NBER/Case Western Reserve Survey of Patentees," NBER working paper No. 7631.

[16] ——, ——, and Henderson, Rebecca , 1993. "Geographic Location of Knowledge Spillovers as Evidenced by Patent Citations," Quarterly Journal of Economics, vol. 108, pp. 577-598.

[17] Levin, R.C., A.K. Klevorick, R.R. Nelson, and S.G. Winter, 1987. "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, vol. 3, 873-820.

[18] Romer, Paul, 1990. "Endogenous Technological Change," *Journal of Political Economy*, vol. 98, pp. 71-102.

[19] Scherer, F.M., 1965. "Firm Size, Market Structure, Opportunity, and the Output of Patented Innovations." *American Economic Review*, vol. 55, 1097-1123.

[20] Scotchmer, Suzanne, 1991. "Standing on the Shoulders of Giants: Cumulative Innovation and the Patent," *Journal of Economic Perspectives*, vol. 5, 29-41.

[21] Shane, Scott, 1999, "Technological Opportunities and New Firm Creation," mimeo, Smith School of Business, University of Maryland.

[22] ——, 2001, "Selling University Technology: Patterns from MIT", mimeo, Smith School of Business, University of Maryland.

[23] Trajtenberg, Manuel, 1990, "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, vol. 21, No. 1.

Table 1: Sample Statistics for Semiconductor Firms:

1979-1995, 636 firm-year observations

	Mean	Median	Min	Max	Std. dev.
Market value ($M)	935.93	114.34	0.0367	48,799.96	3,177.15
Total assets ($M)	630.37	100.13	0.876	18,333.60	1,705.74
Sales ($M)	729.78	110.50	0.1232	16,969.89	1,872.85
Tobin's Q	1.75	1.34	0.085	9.52	1.25
R&D stock ($M)	214.78	26.01	0.0736	4,965.09	593.54
Patent stock	76.61	6.72	0	2,633.78	257.99
Backward citation stock (all obs.)	260.17	19.35	0	9,158.79	859.60
Forward citation stock (all obs.)	4,568.46	244.54	0	177,990.56	15,463.84
D(patent stock = 0)	0.14	0	0	1	0.35
R&D stock/Total assets	0.38	0.27	0.0024	7.56	0.56
R&D stock/Total assets (D(pat > 0))[14]	0.39	0.30	0.0075	7.56	0.59
Backward cites/Total assets	0.48	0.22	0	8.17	0.92
Backward cites/Total assets (D(pat > 0))	0.56	0.29	0	8.71	0.97
Patents/R&D	0.60	0.28	0	19.85	1.34
Patents/R&D (D(pat > 0))	0.70	0.36	0.0025	19.85	1.43
Forward cites/Patents	58.36	24.24	0	2,466.28	151.03
Forward cites/Patents (D(pat > 0))	67.93	32.92	2.27	2,466.28	161.15
Self bwd. cites/total bwd. cites (D(pat > 0))	0.0705	0.0472	0	0.5028	0.0864

[14] Based on 545 firm-year observations whose patent stock is positive. Same below.

Table 2: Correlations Between Different Determinants of Knowledge Stocks: 1979-1995, 636 firm-year observations

	R&D stock	Backward cites	Patents	Forward cites
R&D stock	1.0000	0.8739	0.8337	0.8043
Backward cites		1.0000	0.9459	0.9603
Patents			1.0000	0.8707
Forward cites				1.0000

	R&D/Assets	BCIT/Assets	PAT/R&D	FCIT/PAT
R&D/Assets	1.0000	0.3248	-0.1341	-0.0073
BCIT/Assets		1.0000	0.4744	0.1187
PAT/R&D			1.0000	-0.374
FCIT/PAT				1.0000

Table 3: Estimation of Tobin's q equation

All-firm sample: 636 firm-year observations, 1979 - 1995.

			Citations not deflated		Citations deflated	
	(1)	(2)	(3)	(4)	(5)	(6)
R&D/Assets	0.2597	0.2533	0.2720	0.1683	0.2728	0.1682
	(2.6773)	(2.7896)	(3.0256)	(1.9147)	(3.0244)	(2.0588)
Patents/R&D		0.1339	0.1271	0.1070	0.1231	0.1033
		(1.6971)	(1.7507)	(1.6418)	(1.7561)	(1.5080)
Fwd cites/Patents			0.0030	0.0040	0.0045	0.0037
			(2.0117)	(2.5004)	(1.9565)	(2.8462)
Bwd cites/Assets				0.0753		0.0965
				(3.0120)		(3.6973)
D(Pat = 0)	0.2036	0.1152	0.1345	0.1684	0.1423	0.1447
	(2.4829)	(1.6716)	(1.7222)	(2.1842)	(1.8290)	(1.8744)
LLH	-584.28	-578.53	-561.24	-555.14	-563.93	-558.44
LR test	—	10.60	34.58	12.20	29.20	10.98

Patenting sample: 545 firm-year observations with positive patent stock.

			Citations not deflated		Citations deflated	
	(1)	(2)	(3)	(4)	(5)	(6)
R&D/Assets	0.2587	0.2668	0.2766	0.1752	0.2735	0.1742
	(2.6533)	(2.7477)	(2.8813)	(1.8540)	(2.8371)	(1.9863)
Patents/R&D		0.1459	0.1390	0.0987	0.1368	0.0983
		(1.6732)	(1.7246)	(2.0061)	(1.7295)	(1.9980)
Fwd cites/Patents			0.0037	0.0036	0.0058	0.0053
			(2.3125)	(2.4000)	(2.3200)	(2.5238)
Bwd cites/Assets				0.1176		0.1172
				(2.9772)		(3.6855)
LLH	-489.18	-484.90	-463.93	-458.01	-466.29	-460.52
LR test $\chi^2(1)$	—	8.56	41.94	11.84	37.22	11.54

Note: MLE estimation; t-ratio reported in the parentheses; time dummies included in all equations.

Table 4: Impact of Knowledge Stocks on Tobin's q

	All firms, citations not deflated		Patenting firms, citations deflated	
	Mean	Median	Mean	Median
R&D/Assets	0.38	0.27	0.39	0.30
BCIT/Assets	0.48	0.22	0.56	0.29
PAT/R&D	0.60	0.28	0.70	0.36
FCIT/PAT	58.36	24.24	67.93	32.92
$\frac{\partial \log Q}{\partial (R\&D\ /\ A)}$	0.1075	0.1240	0.1115	0.1334
$\frac{\partial \log Q}{\partial (BCIT\ /\ A)}$	0.0481	0.0555	0.0750	0.0904
$\frac{\partial \log Q}{\partial (PAT\ /\ R\&D)}$	0.0683	0.0788	0.0629	0.0758
$\frac{\partial \log Q}{\partial (FCIT\ /\ PAT)}$	0.0026	0.0029	0.0034	0.0041

Note: calculation based on Table 1 as well as the estimated coefficients in column (4) in the top panel and column (6) in the bottom panel of Table 3.

Table 5: Controlling for Firm characteristics, Patenting Sample

	Citations not deflated					Citation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R&D/Assets	0.1752	0.1392	0.1226	0.1286	0.1235	0.1742	0.1405	0.13
	(1.8540)	(1.9414)	(2.0264)	(1.6572)	(2.1441)	(1.9863)	(1.9873)	(2.98
BCIT/Assets	0.1176	0.1135	0.0978	0.0759	0.4605	0.1172	0.1206	0.11
	(2.9772)	(2.9790)	(2.1733)	(1.4485)	(5.9573)	(3.6855)	(3.0609)	(2.20
interacted with								
log sales					-0.0839			
					(-5.1472)			
Post-82 entrant				0.1095				
				(4.0556)				
Patents/R&D	0.0987	0.0679	0.0478	0.0834	0.0685	0.0983	0.0795	0.05
	(2.0061)	(1.8966)	(1.7903)	(1.2356)	(3.0444)	(1.9980)	(2.3591)	(2.23
Forward cites/Patents	0.0036	0.0054	0.0038	0.0039	0.0041	0.0053	0.0054	0.00
	(2.4010)	(2.3478)	(1.5833)	(2.0526)	(2.1579)	(2.5238)	(2.3478)	(1.95
log sales		0.0918	0.0160				0.0256	-0.02
		(1.9449)	(0.9816)				(1.8028)	(-0.6
Texas Instruments Effect			-0.2515	-0.2132	-0.2017			-0.27
			(-3.2494)	(-4.1398)	(-3.5891)			(-3.5
Post-82 entrant			0.6034		0.6154			0.60
			(9.7796)		(9.4418)			(9.94
LLH	-458.01	-454.52	-443.27	-447.34	-448.66	-460.52	-457.65	-448
χ^2 statistics (LR test)	——	3.49	11.25	10.67	9.35	——	2.87	12.4

Note: MLE estimation; t-ratio reported in the parentheses; time dummies included in all equations.

Table 6: Examining Impacts of Patent Portfolio Size, Patenting Sample

	Citations not deflated					Citatio		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R&D/Assets	0.1752	0.1724	0.1711	0.1747	0.1778	0.1742	0.1770	0.1
	(1.8540)	(2.6401)	(2.7866)	(2.1702)	(2.3456)	(1.9863)	(2.6339)	(2.
BCIT/Assets	0.1176	0.8238	0.5857	0.6075	0.5534	0.1172	0.7582	0.6
	(2.9772)	(2.6370)	(4.6707)	(4.5540)	(4.3782)	(3.6855)	(5.5627)	(4.
interacted with								
log sales				0.0896				
				(0.7232)				
log pat portfolio size		-0.1545	-0.1028	-0.2536	-0.1438		-0.1922	-0.
		(-2.3769)	(-4.7373)	(-1.6839)	(-5.9668)		(-6.6736)	(-4
Post-82 entrant				0.2905	0.2333			
				(1.7660)	(4.0503)			
Patents/R&D	0.0987	0.0858	0.0841	0.0598	0.0637	0.0983	0.0703	0.0
	(2.0061)	(2.4101)	(2.1052)	(2.2411)	(2.3384)	(1.9980)	(2.3049)	(2.
Forward cites/Patents	0.0036	0.0052	0.0035	0.0042	0.0040	0.0053	0.0048	0.0
	(2.4012)	(2.1667)	(1.9441)	(1.9227)	(1.8557)	(2.5238)	(2.4115)	(1.
Texas Instruments Effect		0.1808	0.1336	0.2861	0.2380		0.2520	0.1
		(0.8898)	(1.5924)	(3.0436)	(2.6772)		(2.6087)	(1.
Post-82 entrant			0.5168					0.5
			(8.6711)					(8.
LLH	-458.01	-446.56	-441.15	-437.33	-439.77	-460.52	-446.68	-44
χ^2 statistics (LR test)	——	11.45	16.86	20.68	18.24	——	13.84	17.

Note: MLE estimation; t-ratio reported in the parentheses; time dummies included in all equations.

Table 7: Average Value of Backward Citations and Patent Portfolio Sizes, Patenting Sample, Citations Stock Deflated

	Patent portfolio size	Column (8), Table 6		Column (10), Table 6			
				Pre-82 firms		Post-82 firms	
		$\widehat{b}_2$	$\widehat{b}_2/\widehat{b}_1$	$\widehat{b}_2$	$\widehat{b}_2/\widehat{b}_1$	$\widehat{b}_2$	$\widehat{b}_2/\widehat{b}_1$
Lower 25%	8	0.3483	2.08	0.2831	1.51	0.5335	2.84
Median	28	0.1887	1.13	0.0688	0.37	0.3192	1.70
Top 25%	95	0.0330	0.20	-0.1403	-0.75	0.1101	0.59

Note: calculation based on estimates in columns (8) and (10) of Table 6.

Table 8: Spillovers Within and Beyond the Firm and Industry: Non-semiconductor, External Semiconductor, and Self Citations, Patenting Sample, Citations Deflated

	(1)		(2)		(3)		(4)	
R&D/Assets	0.1807	(2.0418)	0.1779	(2.0262)	0.1857	(2.0519)	0.1742	(1.9931)
BCIT/Assets	0.2418	(2.2640)	0.1685	(2.7944)	0.6660	(3.7374)	0.9081	(4.3429)
interacted with								
log pat portfolio size					-0.1763	(-6.5784)	-0.2096	(-5.3606)
Post-82 entrant					0.3793	(5.2901)	0.1297	(1.5081)
NSCBCIT/Assets	-0.0602	(-0.3472)	-0.0748	(-1.2026)	-0.1048	(-1.6375)	-0.0957	(-1.9732)
interacted with								
log pat portfolio size								
Post-82 entrant								
SelfBCIT/BCIT	0.1475	(1.4647)	0.6101	(2.1970)	0.6851	(2.2776)	-0.7785	(2.3527)
interacted with								
log pat portfolio size							0.2660	(1.0870)
Post-82 entrant							3.9619	(5.5388)
Patents/R&D	0.0705	(3.0388)	0.0685	(2.7183)	0.0668	(2.1974)	0.0743	(2.4603)
Forward cites/Patents	0.0055	(2.1154)	0.0041	(2.1579)	0.0047	(2.1364)	0.0038	(2.7143)
Texas Instruments Effect			-0.2534	(-4.6325)	0.2864	(2.8413)	0.2130	(1.3777)
Post-82 entrant			0.6078	(10.4433)				

Note: MLE estimation; t-ratio reported in the parentheses; time dummies included in all equations.

Appendix

Construction of Tobin's Q and R&D stocks from Compustat

We extract the market and book value as well as R&D expenditures of the 120 semiconductor firms from Compustat database, from 1979 to 1995. Tobin's Q is defined as

$$Q = \frac{MKVALM + DT + PSTK}{AT} \tag{A.1}$$

where $MKVALM$ is the "sum of all the company's trading issues multiplied by their respective monthly closing price" by the end of each year, DT refers to the amount of total debt including both the long-term and short-term debt, $PSTK$ is the market value of preferred shares of the company, and AT represents the "current assets plus net property, plant and equipment plus other noncurrent assets." All monetary values are adjusted for inflation based on U.S. GDP deflator and are in units of million 1998 U.S. dollars.

The R&D capital stock is constructed as the accumulated current and past R&D expenditures, assuming an annual depreciation rate of 15%. As we do not have data on R&D expenditures before 1979, we assume them to be zero. This unambiguously leads to a *under*-estimation of the R&D stocks for firms actively engaged in R&D activities before 1979. However, only 18 semiconductor firms in this sample had nonzero R&D expenditures in 1979, and only 4 of them had R&D expenditures more than 6 million dollars in 1979 (Advanced Micro Devices, Intel, National Semiconductor Corp., and Texas Instruments), implying that this downward bias should be very small.

Truncation of Citation Counts and Citation Inflation

To deal with the data truncation problem of forward citations, we follow HJT (2005) and estimate a structural citation-lag model. In particular, it is assumed that the fraction of lifetime forward citations in each year after the initial patent application follows a stationary double-exponential distribution and is independent of the overall lifetime citation intensity, and the frequency of a cohort t patent being cited by a cohort $t + s$ patent is

$$c_{t,t+s} = \beta_0 \alpha_t \gamma_{t+s} \exp(-\beta_1 s)(1 - \exp(-\beta_2 s)) \tag{A.2}$$

where β_0 measures the overall citation intensity, s denotes the citation time lag, $\exp(-\beta_1 s)$ describes a diffusion process and $(1-\exp(-\beta_2 s))$ characterizes an obsolescence process (Jaffe, and Trajtenberg (1996)). α_t and γ_{t+s} are two time dummies for cited and citing year, respectively.

In this study we further distinguish between citations occurred within the same firm (self citations), beyond the firm but still in the same narrowly defined technological field of semiconductor, and citations made by patents from different technological fields. We make such distinctions out of concern that knowledge flows may occur at different speed in these cases. Thus, we formulate the following estimation equation

$$\log(c_{t,t+s,j}) = \log(\beta_0^j) + \log(\alpha_t) + \log(\gamma_{t+s}) - \beta_1^j s + \log(1 - \exp(-\beta_2^j s)) + \varepsilon_{t,t+s,j} \tag{A.3}$$

where $c_{t,t+s,j}$ is the frequency of a cohort t patent being cited by a cohort $t+s$, and j indicates whether the citation occurs within the same firm, from a different firm but within the same technological field of semiconductor, or from a different firm and in a different technological field.

Equation (A.3) is estimated using maximum likelihood, assuming $\varepsilon_{t,t+s,j}$ is *i.i.d.*, normally distributed. Based on the model estimation we can then construct the model-implied citation frequency in the years observed in the dataset, net of time dummies and overall citation intensity, as

$$D_{t,1996,j} = \sum_{s=1}^{1996-t} \exp(-\beta_1^j s)(1 - \exp(-\beta_2^j s)) \tag{A.4}$$

where 1996 is the last year of citation records that we use (citations from cohorts 1997 to 1999 are incomplete in the database because many of those patents had not been granted by the end of 1999). The citation frequency for years not observed in the database, conditional on the citations observed in the database, can then be projected as

$$c_{t,t+s,j} = \frac{N_{t,1996,j}}{D_{t,1996,j}} \exp(-\beta_1^j s)(1 - \exp(-\beta_2^j s)) \tag{A.5}$$

where $N_{t,1996,j}$ is the sum of actual number of forward citations observed in the database.

For the late 1980s and 1990s patents, there is an additional problem: because the forward citations are often zero in the first several years, $N_{t,1996,j}$ could be zero, so equation (A.5) will project zero lifetime citations for them. However, citation counts are bounded below by zero, and the expected number of lifetime citations should be positive. Thus in such cases we use the empirical expectation of citations observed in the first 20 years after patent applications, conditional on observing zero citations in the first M years, $M = 1, 2, .., 10$:

$$E\{\sum_{j=0}^{20} N_{t,t+j} | \sum_{j=0}^{M} N_{t,t+j} = 0\} \tag{A.6}$$

as the prediction of total citations that will be observed in the first 20 years for those patents. Specifically, we estimate the empirical expectation in equation (A.6) for cohorts 1975 to 1978, for which we have an actual 20 years of citation observations in the citations database, and assume that is the expected total citations any patent in cohort 1986 to 1995 will receive in their first 20 years, conditional on they had received zero citations by 1996.