

Watzinger, Martin; Schnitzer, Monika

**Conference Paper**

## Measuring Spillovers of Venture Capital

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik  
- Session: R&D and Innovation, No. A03-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Watzinger, Martin; Schnitzer, Monika (2014) : Measuring Spillovers of Venture Capital, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: R&D and Innovation, No. A03-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/100318>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Measuring Spillovers of Venture Capital

Monika Schnitzer      Martin Watzinger\*

1st March 2014

In this paper we measure knowledge spillovers arising from venture-capital financed companies on the patenting activities of other companies and compare them to spillovers from established firms. We develop a novel measure to identify the appropriate spillover pool based on backward citations which reflect channels for learning between firms. Using panel data of U.S. firms we show that venture capital investment in start-ups generates significant spillovers on the patent quantity and quality of other firms. Counterfactual estimates suggest that these spillovers are larger than those generated by corporate R&D. We address potential concerns about causality with an instrumental variable strategy using changes in federal and state tax incentives as instrument for R&D and past fund raising as instrument for venture capital investment.

---

\*University of Munich, Department of Economics, Akademiestraße 1, 80799 Munich, Germany; Monika Schnitzer: Schnitzer@lrz.uni-muenchen.de; Martin Watzinger: martin.watzinger@lrz.uni-muenchen.de. We thank John van Reenen and Nicholas Bloom for sharing their data on federal and state tax credits. Financial support of the Deutsche Forschungsgemeinschaft through SFB-TR 15 is gratefully acknowledged.

# 1. Introduction

Governments around the world are eager to stimulate the venture capital industry through public policy: venture capital funds are tax free in France and the UK, the Canadian government directly acts as a venture capitalist through the Business Development Bank of Canada and the European Union provides financing for venture capital funds with the help of the European Investment Fund. One possible reason for government intervention is that VC financed start-ups contribute to innovation: Between 1983 and 1992, venture capital financed R&D accounted for 8% of industrial innovation in the US, while the ratio of venture capital to R&D was only 3% (Kortum and Lerner, 2000). Innovation in itself is not a market failure, yet a large body of literature showed that corporate R&D exhibits large positive spillovers and consequently the observed investment in R&D is lower than the social optimal amount (Bloom, Schankerman, and Van Reenen, 2013).

There are at least three reasons to doubt, that spillovers of venture capital are similarly large than the spillovers of corporate R&D and therefore deserve a preferential treatment by the government: First, venture capital is used for all kind of investments such as advertising and not only for investment in research. Second, the aim venture capital financed start-ups is to bring a particular product to the market and therefore their research might be more narrow in focus than the research of corporate R&D. Third, Bloom, Schankerman, and Van Reenen (2013) find that small firms generate few spillovers, as they tend to operate more in technology niches. As venture capital backed start-ups are naturally small this might point to a limited potential for spillovers.

In this paper we measure knowledge spillovers of venture capital-backed start-up companies and compare their relative size to knowledge spillovers to the external effects of corporate R&D. We find that VC-financed firms generate positive spillovers on other firms' quantity and quality of patents. Counterfactual simulation suggest that per dollar of investment knowledge spillovers from venture capital are significantly larger than spillovers from conventional R&D. Descriptive evidence points to two potential reasons: First, research of venture capital financed start-ups is particularly innovative measured by the number of cites and the generality of the resulting patents. Second, start-ups do not commercialize technological niche products, but in contrast are central in technology space.

For this study, we follow a large literature assuming that higher spillovers increase the productivity of every dollar invested in innovation. The reason is that knowledge "spills" over and thus learning between companies occurs. If we observe that venture capital-financed R&D results in larger spillovers, i.e. leads to a larger increase in research output

of a given company for a given investment than conventional R&D does, then we can conclude that the external effect of start-ups is larger than that of established companies. The innovation output of a company can be readily measured by the quantity and quality of patents and the input can be gauged by the size of the R&D stock or the total amount of venture capital received of a company. The problem is, however, that there is no direct measure for the size of spillovers.

To measure spillovers, the crucial challenge is to identify the “spillover pool”, companies from which a particular company might learn and therefore spillovers could originate. However, as knowledge flows are in general unobserved, identifying this spillover pool has proven to be challenging. One often used approach is to include companies that are close in physical space, the idea being that knowledge is communicated via personal contacts, giving rise to local spillovers. Another approach originated by Jaffe (1986) is to include in the spillover pool companies which use the same technologies measured by the technology classes of the their patents and which therefore are close in technology space. The rationale is that companies which work for example on “robots” (technology class 901) learn most from other companies working on robots. Bloom, Schankerman, and Van Reenen (2013) extended this concept by also including companies which work in “similar” technologies and call this new measure Mahalanobis distance. Two technologies are similar if companies often hold patents of the two corresponding patent classes together in their patent portfolio. For example patents about robotics and artificial intelligence are complementary and therefore collocated in companies.

While both approaches to define the spillover pool, geographical distance and technological distance, are intuitively appealing, they are based on an a priori reasoning about the learning process and not on direct evidence of knowledge flows. This conceptual weakness was first recognized by Jaffe, Trajtenberg, and Henderson (1993). They showed that patents cite more often patents of companies which are geographically closer than of companies that are at a larger distance. In the innovation literature backward citations between patents are an accepted direct - albeit noisy - measure of knowledge flows.<sup>1</sup> Thus, this observation vindicates the use of geographical distance to define the spillover pool. In this paper, we suggest to go one step further and to use the direct information on knowledge flows embodied in the backward citations to define the spillover pool of a company. For example, suppose patents about robotics often cite patents about information processing but not vice versa. Then the spillover pool of robotics companies should include companies in semiconductors and information processing but not the other way

---

<sup>1</sup>Jaffe, Trajtenberg, and Fogarty (2000) showed with a survey that if a patent cites another there was indeed a direct interaction between the inventors in 60% of all cases.

round.

To implement this idea, we develop in the spirit of Bloom, Schankerman, and Van Reenen (2013) a weighing matrix based on the citation propensities between different technology classes to modify the technological distance measure of Jaffe (1986). The advantage of this “citation-augmented” distance measure is two-fold. (i) It allows to capture knowledge flows between companies that are not necessarily close in technology space, but that have a (backward citation) proven record of learning from each other. Thus, it accommodates the insight that important sources of knowledge spillovers lie outside the industry in which a firm operates. (ii) In contrast to the Jaffe- and the Mahalanobis-measure, it allows spillover flows between companies to be asymmetric (as backward citations between two technology classes can be asymmetric) while by construction the technology distance never is. We think this measure to be both a more general and a more realistic specification of knowledge flows among firms that are certainly anything but symmetric.

Our data come from two sources, Compustat and VentureXpert. Compustat contains balance sheet data for all U.S. publicly listed companies. VentureXpert is a prime source for venture capital investment and fund-raising data. We select all companies which patented at least once in the period from 1990 to 2005. Patent data are from the NBER U.S. Patent Citations Data File and from data files of Lai, Amour, Yu, Sun, Torvik, and Fleming (2011) for which we create a name match per hand to the venture capital data. For the Compustat Data, the NBER provides a unique identifier to match the balance sheet data with patent counts and cites.

Using our new distance metric we calculate for each company in Compustat and in VentureXpert a value for the spillover measure for established companies and for venture capital financed companies, respectively. In our estimation we then use the two new spillover measures as explanatory variables for the patenting productivity of a company along with its R&D or venture capital stock. A major challenge in the estimation procedure is that a company does not randomly invest in innovation but changes its outlays responding to its own productivity and the technological opportunities available. To see the potential endogeneity problem, suppose a particular technology field experiences a positive technological shock. Then it might be easier to produce patents in this field and at the same time companies working in the same technology field might increase their research and development outlays. Similarly, venture capital funds might allocate their investment to technology fields hit by a positive shock. Therefore the two spillover terms might pick up the effect of technological progress and as a result could be spuriously high. To address this endogeneity problem we use an instrumental variable approach with two instrumental variables.

We instrument R&D expenditures of established companies with the level of R&D tax credit in a state as in Bloom, Schankerman, and Van Reenen (2013) and venture capital investment with past fund-raising of buyout funds (Nanda and Rhodes-Kropf, 2012). The idea is that the introduction of R&D tax credits in the different U.S. states has a direct influence on the level of research and development by lowering costs. At the same time, it is unlikely that government officials are able to react in time to a change in the technological frontier. Venture capital spending is instrumented with past fund-raising of private equity buyout funds. Buyout funds and venture capital funds belong both to the class of private equity. Institutional investors often allocate funds to private equity without distinguishing between the two subclasses. Therefore buyout fund-raising is correlated with venture capital fund-raising but potentially uncorrelated with the arrival of technological opportunities of VC backed companies.

Our results show that not only established firms, but also VC-financed firms generate significantly positive spillovers on other firms' quantity and quality of patents. We find consistently strong spillovers for established firms for all alternative definitions of the spillover pool. In contrast, for VC-financed start-ups, the results differ across different distance measures. If we use the information on backward citations to construct the spillover pool, we find that the spillover effects of VC-financed firms are statistically significant and positive. In contrast, using the Jaffe or the Mahalanobis measures there are only little measurable spillover effects. This provides evidence in support of our conjecture that using backward citation patterns to define the spillover pools improves the measurement of spillovers.

We also provide a back of the envelope calculation of the relative size of the spillover effects. We consider two counterfactual scenarios, one in which we increase R&D spending by 1 Million US \$, and one in which we increase R&D spending by 10 percent. We find for both scenarios that the external effect of R&D and venture capital investment is at least 10% of the social return and that the relative share of external versus private returns is more than 50% larger for VC-financed firms than for established companies. Taking the result of our most conservative specification at face value, a one million US \$ increase in R&D spending results on average in 0.01 patents for other companies while one million US \$ more venture capital investment results in 0.13 patents for other companies.

A reason for the large spillovers of venture capital might be that venture capital financed companies are central in technology space. This stands in contrast to the findings of Bloom, Schankerman, and Van Reenen (2013) which find that small companies are in technological niches and therefore lack the potential to strongly influence other companies. In addition research of venture-capital financed start-ups also seems to be particular

innovative: Patents of these companies are highly cited and more general than the average. Yet, this is true even in the year in which these companies receive their first investment. This points to venture capitalist choosing to fund companies with a large spillover potential and not to investors changing the effectiveness or the focus of the research undertaken.

The contribution of our paper is thus twofold. First, we develop a novel distance measure between firms based on patent citations instead of technology distance alone. This allows us to better capture the knowledge flows generated from VC-financed firms which tend to be small and hence almost by definition not likely to generate spillovers if the Jaffe or Mahalanobis index is used. Second, we are the first to provide a direct measurement of innovation spillovers generated by venture capital financed firms and to compare the size of these spillovers to spillovers of corporate R&D. Our results confirm that VC-financed firms generate positive spillovers and that these spillovers are relatively larger than those of established companies' R&D.

Our results complement the findings of Kortum and Lerner (2000). In their seminal paper, Kortum and Lerner find that venture capital financed R&D is much more potent in producing patents than R&D financed by established companies. However, as their analysis is based on aggregated industry data, it does not allow to disentangle direct and indirect effects of venture capital financing on patent production. Thus, while it shows that VC is overall very effective in stimulating patent production, it does not provide evidence on how much of this direct effect is due to knowledge spillovers, let alone on how large such spillovers are in comparison to spillovers generated by conventionally financed R&D. This study also contributes to the large literature measuring the private and social returns of R&D undertaken by established companies which is summarized in Hall, Mairesse, and Mohnen (2009). In general, this literature finds large and positive social returns of R&D. However, it does not yet consider any differential effects from venture capital financed research.

The paper proceeds as follows: In section 2 we describe the data used and provide summary statistics. In section 3, we lay out the empirical strategy for measuring spillovers of R&D and venture capital. In section 4 we present our empirical results and section 5 concludes.

## 2. Data and Descriptive Statistics

For our dataset we combine patent data with firm level data of venture capital financed companies and established companies in the US from 1990 to 2005. The patent data is from the NBER containing all utility patents filed in the US with name of the applicant, year of application, the state of application, the number of cites a patent receives and a classification according to the 3-digit US patent class. This classification is determined by the patent examiner and sorts patents into one of 400 functional groups.<sup>2</sup> This classification gives a fine grained view of the wide variety of patents and correspondingly of the technological focus of the patent assignee. To complement the data at the end of the considered time period we use the Harvard patent dataset of Lai, Amour, Yu, Sun, Torvik, and Fleming (2011). The resulting dataset contains around 2.45 million patents and is matched to the two firm data sources which we discuss next.

Our firm level data source for established companies is the US Compustat file. This data can be easily matched to the the NBER Patent-Citation Data File using a unique identifier provided by the NBER (Hall, Jaffe, and Trajtenberg, 2001). The Compustat File contains yearly accounting data for US publicly listed companies with company name, the fiscal year, the state of the firm head quarter, the four-digit SIC code, sales and research and development expenditures. To calculate the R&D stock from R&D expenditures we apply the perpetual inventory method with a 15% depreciation rate (following inter alia Hall, Jaffe, and Trajtenberg 2005). Additionally, we delete all companies whose data we conceive as erroneous: for example companies whose sales increase by more than 100% or companies which produce patents with zero R&D spending. The resulting database contains 1317 companies with 232'846 patents.

The second firm level data source is Thomson Reuters VentureXpert which comprises investments data in US venture capital (VC) financed companies. Each record contains the name of the investee company, the investment date, a four-digit SIC code and the estimated amount invested. The latter is our main measure for total available funds of a company. In addition we know the investment stage of the company, i.e. if it is a seed, an early stage, expansion or late stage investment. We delete all companies whose first investment is before 1990. Analogous to the calculation of the R&D stock we calculate the VC stock with perpetual inventory method.

We match this investment data by company name to the patent data with the help of algorithms from the Apache Lucene library and check the results by hand. The start-up data set contains 1672 companies with 12'791 patents. Summary statistics for the dataset

---

<sup>2</sup>In theory there are 800 functional groups, yet we only observe a positive patent count in 400 of them.



are shown in Table 1.

Table 1: Summary statistic

<b>Established Companies</b>						
	mean	sd	min	max	p10	p90
Average R&D in million dollar	42.81	127.90	0	1819	1	93
Average Sales in million dollar	1176.77	4892.96	0	125759	13	2506
Number of patents (total)	176.80	666.62	0	12383	1	371
Forward cites (total)	1965.90	8310.43	0	171352	8	3770
Generality (total)	49.83	194.34	0	3806	0	97
Years in the data	10.97	4.18	1	16	5	16
Number of companies	1317					
<b>Venture Capital-backed Companies</b>						
	mean	sd	min	max	p10	p90
Total funding in million dollar	2.29	4.32	0	41	0	8
Number of patents (total)	7.65	12.54	0	136	0	20
Forward cites (total)	165.02	446.62	0	10048	0	400
Generality (total)	2.82	5.07	0	72	0	7
Years in the data	5.09	2.88	1	16	2	9
Number of companies	1672					

### 3. Empirical Strategy

In this section we discuss first the measures used for innovative activity and then different ways to define the spillover pool which may generate spillovers.

#### 3.1. Using Patenting Activity to Measure Innovation

Measuring the innovative activities of companies is intrinsically difficult. The challenge is in particular to find measures that are available for all the firms that are in the focus of the investigation, not just for subsamples of firms. In the innovation literature, it is generally accepted that patenting activity of firms is a suitable output measure. The advantage of using patent data for our analysis is that they are available for both established as well as for VC-finance companies. Moreover, using patent related measures allows us to capture not only the quantity of firm innovation activities, but also their quality.

For our analysis, we follow the practice of Hall, Jaffe, and Trajtenberg (2001) and use several variables to measure patenting quantity and quality: patent count, patent cites, and generality. Patent count is a purely quantitative measure of the extent of

innovative activity. Patent cites counts the number of citations a patent receives and thus reflects the quality of the innovation (Hall, Jaffe, and Trajtenberg, 2005). Furthermore, to measure the general applicability of a patent we calculate for each patent the Herfindahl index across technology classes for the patents by which the patent is cited (Trajtenberg, Henderson, and Jaffe, 1997). Thus we capture the dispersion across technology classes of patents using the patent. The “generality” measure used in the regression is calculated as one minus this Herfindahl Index. One potential concern to be taken into account is that the citing behavior might change from year to year and from technology to technology. To account for these changes, we scale the last two measures by the average value in a particular year and a technology class following Bernstein (2012).

### 3.2. Definition of the Spillover Pool: Distance Measures

The rationale for knowledge spillovers is that a firm may learn from other firms. One reason might be that scientists from both companies meet and exchange ideas, or a firm hires scientists previously employed by another firm, or a firm learns via scientific publication or by reverse engineering. The canonical way to establish innovation spillover effects is to identify which of the companies are most likely to generate spillovers, define them as the spillover pool and find an empirical association of R&D spending of the spillover pool with outcome variables of the firms under consideration. Obviously the main challenge is to identify which of the companies a firm is most likely to learn from as learning is in general unobserved.

In the literature, the two most common approaches are based on geographical and technological distance. The first of these two approaches is based on the idea that knowledge is communicated via personal contacts. Thus, the idea is to focus on companies that are close in physical space and hence give rise to local spillovers (Jaffe, Trajtenberg, and Henderson, 1993). An alternative definition of a spillover pool has been developed by Jaffe (1986). According to his definition, the spillover pool of a particular company consists of companies that are close in the “technology space”. The rationale here is that companies learn from other companies that are active in same technologies. The Jaffe distance measure is the uncentered correlation between the patent share vectors

$$Tech_{i,j}^{Jaffe} = s'_i s_j$$

where  $s_i$  is the share of patents of company  $i$  over various technology classes standardized by their firm patent share dot product  $s_i = \frac{S_i}{(S_i S'_i)^{\frac{1}{2}}}$ .

Bloom, Schankerman, and Van Reenen (2013) extended this concept, using a Maha-

lanobis distance metric. Their definition of a spillover pool is based on the idea that companies learn not only from companies using the same technology but also from companies using similar technologies. The similarity of two technologies is measured by how often these patents embodying these two technologies are collocated in a company. To implement this concept, Bloom, Schankerman, and Van Reenen (2013) develop the Mahalanobis distance measure introducing a weighing matrix  $\Omega^{Mal}$  to the Jaffe distance measure, arriving at

$$Tech_{i,j}^{Malhanobis} = s_i' \Omega^{Mal} s_j$$

The quality of these alternative definitions of the spillover pool depends on how well they capture the learning of companies from other firms. While both types of measures, geographical distance and technological distance, are appealing in indicating a potential for spillovers, a drawback is that they are not based on direct evidence for actual learning to happen.

A direct measure for such learning is given instead by backward citations on patents. Patent A cites patent B if the patent B constitutes “prior art”, i.e. a backward citation indicates that patent A draws on the knowledge of patent B. As Jaffe, Trajtenberg, and Fogarty (2000) show, backward cites do indeed constitute a valid measure for knowledge flows. According to them, around 60% of all backward cites correspond to some form of interaction between the inventors.<sup>3</sup> Deng (2008) estimates the economic value of the knowledge spillover derived from a backward citation in the US semiconductor industry in the 1980s and 1990s to be on average between \$ 0.6 and 1.2 million R&D equivalent dollars. An indication how these knowledge flows may depend on geographical distance is provided by Jaffe, Trajtenberg, and Henderson (1993) who show that companies are more likely to cite other companies which are more closely in physical space. As Griffith, Lee, and Van Reenen (2011) show this home bias of knowledge spillovers has fallen over time, more so in traditional sectors and less so in pharmaceuticals and IT and communication technologies. However, it has not been analyzed yet to what extent companies are more prone to cite companies which are close in technology space. To study this question we plot in Figure 1 the cross citation patterns between relatively broad technology subcategories defined in Hall, Jaffe, and Trajtenberg (2001). For visualization we use the broad technology subcategories because they are easier to interpret. Every technology subcategory comprises several technology classes and the mapping is given in the appendix of Hall, Jaffe, and Trajtenberg (2001).

---

<sup>3</sup>Nelson (2009) compares different measures of technology diffusion such as backward citations, licensing and publications. He concludes that patent citations are the most restrictive measure of diffusion which suggests that the spillover effects identified provide a lower bound of the true effects.

Figure 1: Cross-Citation Patterns

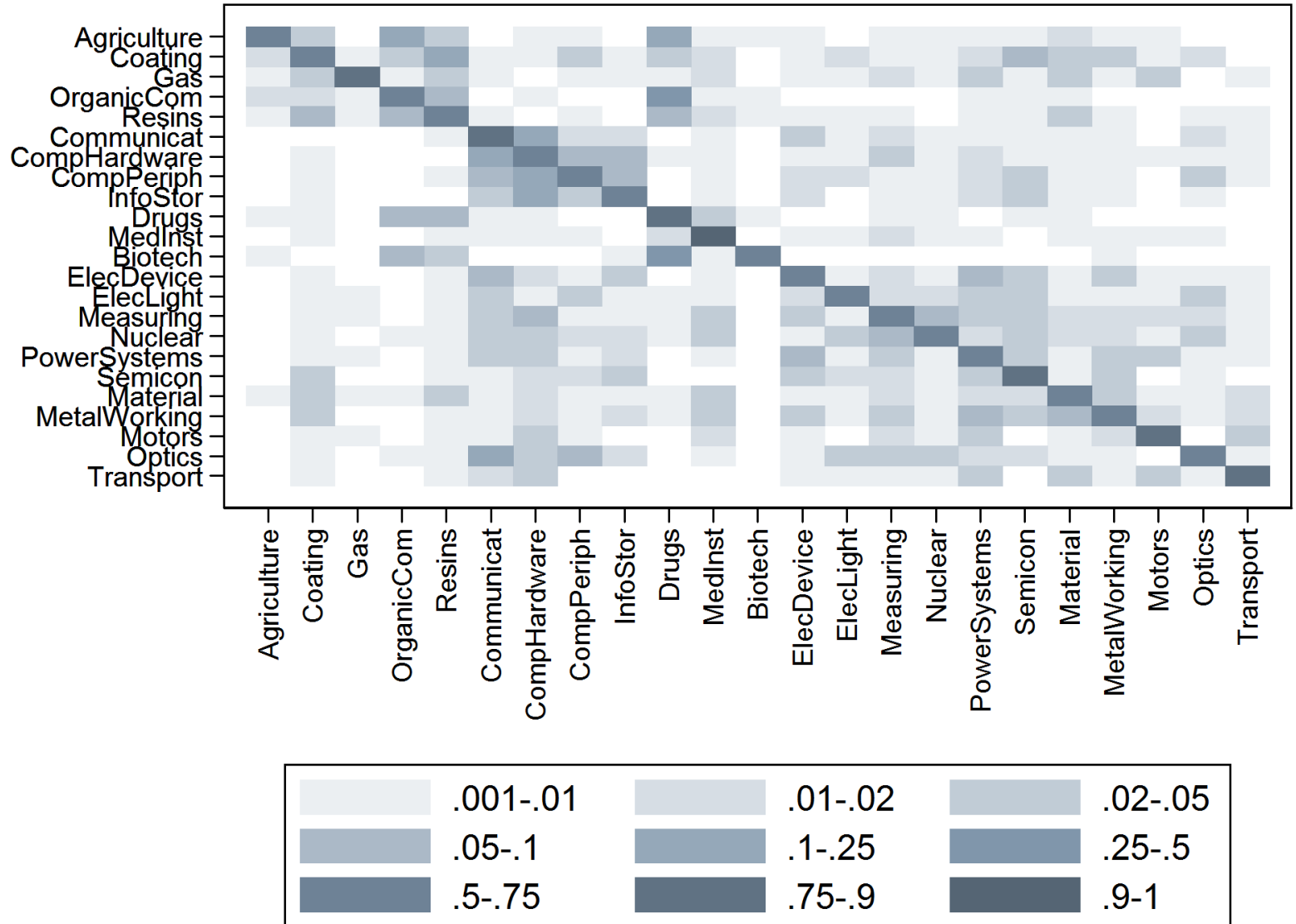


Figure 1 depicts how many backward citations (in shares) patents in a technology subcategory on the y-axis draw from patents in a technology subcategory on the x-axis in the year 2000. Apparently, there is indeed a strong tendency of patents in a particular technology class to cite patents from their own technology subcategory. Interestingly, however, there is also a significant number of citations from other technology subcategories. One can visually see “clusters” of backward citations between similar technologies, as for example in computer hardware and software, comprising communication, computer periphery, and information storage. Another such cluster is drugs with organic compounds, resins, medical instruments and coating.

Figure 2 illustrate these technology clusters. As expected, almost three quarters of the backward citations used in patents from the technology class Drugs are taken from the technology class Drugs itself. Another 9% of its backward cites, however, are from organic compounds. Similarly, patents in the technology class Hardware and Software draws 60% of its backward cites from its own technology class. But more than 15% of its backward cites are drawn from communications and another 10 % from the technology class “Information Storage”.

Figure 3 illustrates the share of backward cites a patent in a particular technology subcategory draws from technologies classes other than its own. This share is on average around 30%. A particularly low share of outside citations are medical instruments that draw only 10% of its backward cites from outside its own technology class, while in contrast coating draws almost 50% of its backward cites from outside.

Another interesting observation to be made from Figure 1 is that the pattern of backward citations is not symmetric. While patents in the agricultural technology class cites patents from drugs, drug patents do not cite patents from agriculture.

Taken together, this evidence suggests that there are significant knowledge flows across technology classes, although not necessarily symmetric ones; Companies learn not only from companies which do the same thing technological wise but also from companies which work in different technology fields.

We suggest therefore to use the information provided by backward citations and incorporate the citation pattern in defining the spillover pool. In particular, we develop a weighing matrix based on the backward citations and use this weighing matrix to modify the Jaffe-distance measure. The advantage of this citation-augmented distance measure is two-fold. (i) It allows to capture knowledge flows that occur between companies that are not necessarily close in technology space, but that have a (backward citation) proven record of learning from each other. (ii) In contrast to the Jaffe- and the Mahalanobis-measure, it allows spillover flows between companies to be asymmetric, as

Figure 2: Citation patters from other technologies

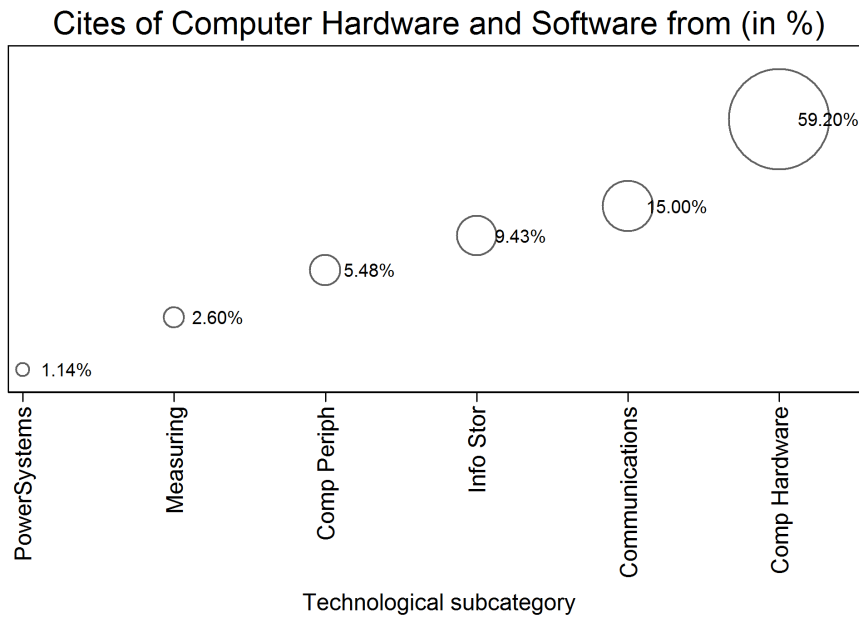
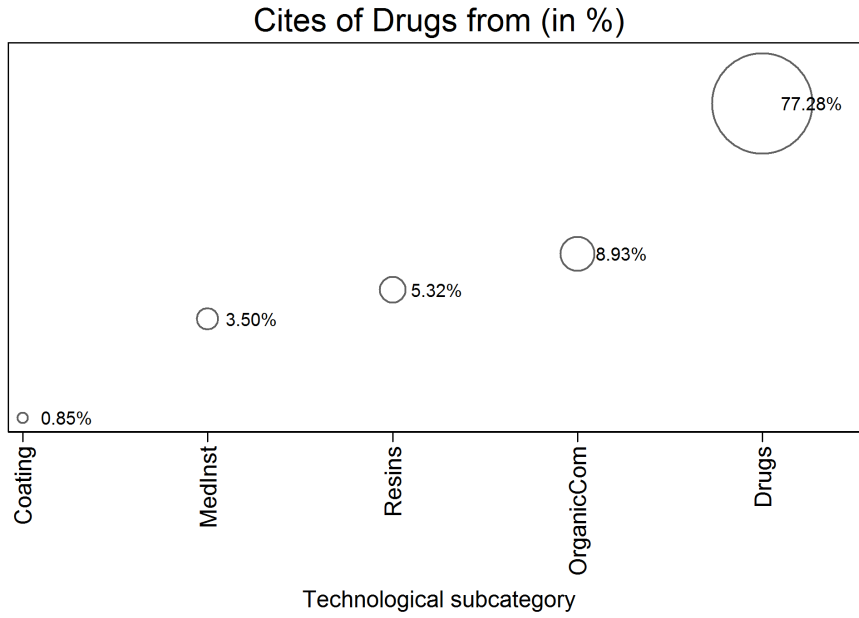
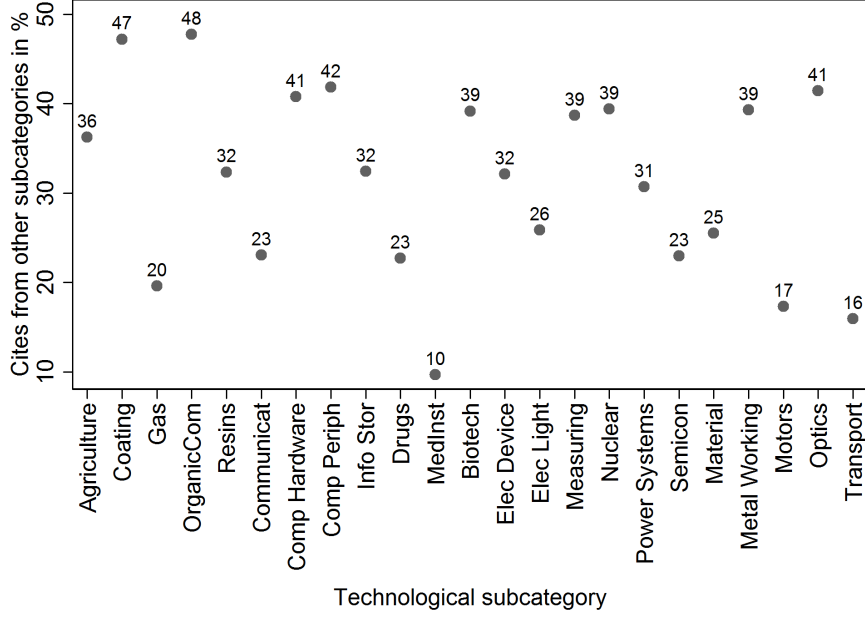


Figure 3: Share of cites from other technological subcategories



backward citations between two companies can be asymmetric while by construction the technology distance never is.

More specifically, we incorporate the citation pattern in the distance metric using a weighing matrix following Bloom, Schankerman, and Van Reenen (2013). To define the distance between technology class  $k$  and  $l$  we calculate the share of cites  $l$  draws from  $k$ , i.e.

$$\omega_{l,k} = \frac{Cites_{l,k}}{\sum_m Cites_{l,m}}$$

Then we arrange these weights in a matrix  $\Omega^{Cites}$ . Note that  $\Omega^{Cites}$  is asymmetric because it traces the flow of knowledge between technology classes.

The resulting citation-augmented distance between company  $i$  and company  $j$  is then given by

$$Tech_{i,j}^{Cites} = s_i' \Omega^{Cites} s_j$$

A simple example shows that using this citation-augmented distance measure, we can trace knowledge flows that the Jaffe index would fail to capture. Suppose Company D has all its patents in the technology “Drugs” and company A has all its patents in “Agriculture”. Then the patent share vectors are  $S_D = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  and  $S_A = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ . As

there are no common technologies, the Jaffe distance measure would yield a value of 0, indicating that no spillovers occur between these two companies. Suppose now that 50% of all cites from agriculture are from drugs and 50 % from itself, while drugs only cites itself. This leads to the following citation matrix:

$$W^{Cites} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \end{bmatrix}$$

The citation-augmented distance of the agricultural company A to the drugs company D is hence

$$P_{AD}^{Cites} = S_A' W^{Cites} S_D = 0.5.$$

However, D does not get any knowledge from A, so

$$P_{DA}^{Cites} = S_D' W^{Cites} S_A = 0.$$

Figure 4 plots the different distance metrics for our sample. This plot shows that there is independent information in our citation-augmented distance measure as compared to both, the Jaffe and the Mahalanobis measure. For instance, while the Mahalanobis weight only increases distance relative to the Jaffe measure our citation-augmented distance can both increase and decrease the measure relative to the Jaffe measure.

As indicated above, the citation-augmented distance measure is markedly asymmetric. In Figure 5 we plot the distance between companies pairwise, i.e. the distance from company 1 to company 2 at the vertical and the distance from company 2 to company 1 at the horizontal axis for 0.5% sample of our data. The distances are of course positively correlated, but not perfectly so.

### 3.3. Empirical Design

We are interested in estimating the following log-linearized Cobb-Douglas type production function of company  $i$  at time  $t$ :



Figure 4: Plotting different distance measures against each other

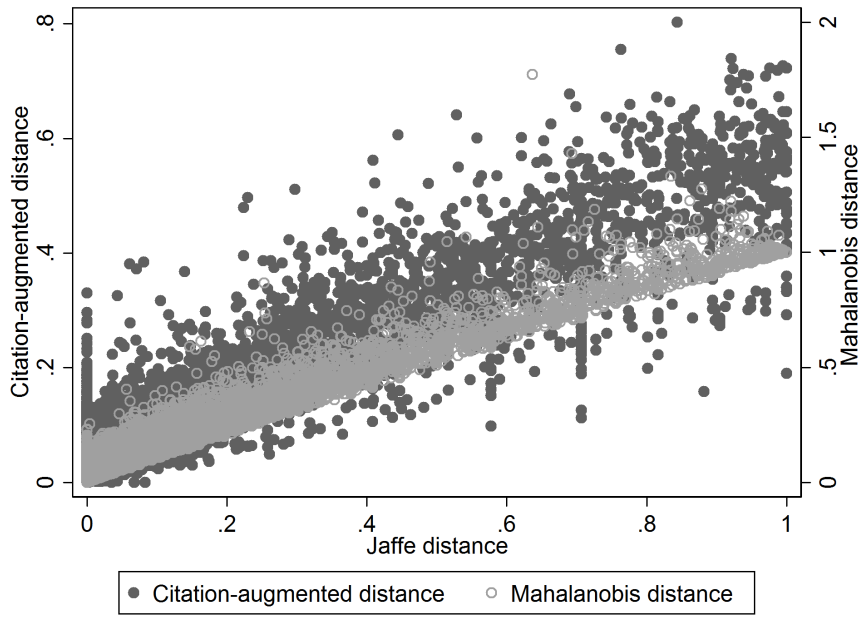
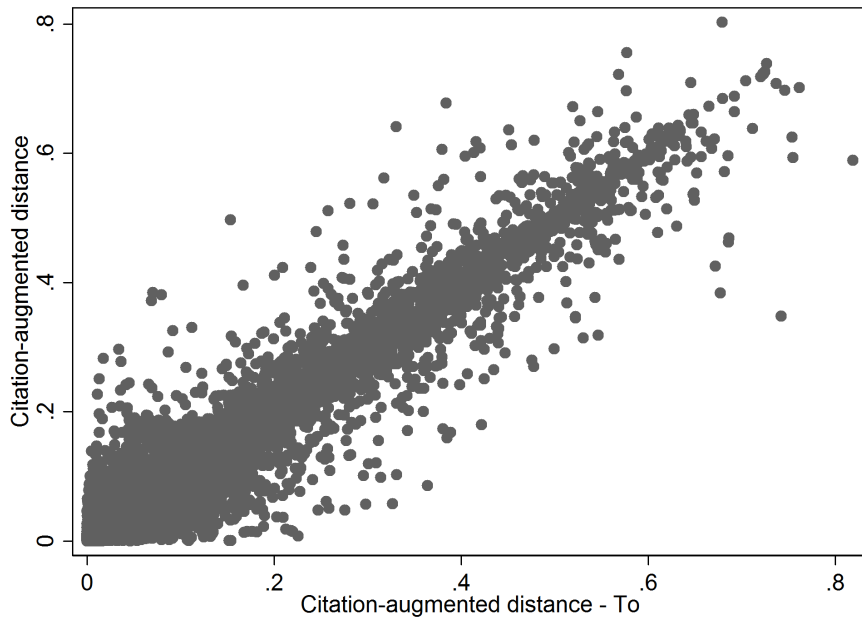


Figure 5: Citation-augmented distance from and to a company



$$\begin{aligned}
Outcome_{i,t} = & \beta_0 + \beta_2 \cdot \ln \left[ \left( \sum_{j \neq i} Tech_{i,j}^{Cites} \cdot R\&D_{i,t-1} \right) + 1 \right] \\
& + \beta_4 \cdot \ln \left[ \left( \sum_{j \neq i} Tech_{i,j}^{Cites} \cdot VC_{j,t-1} \right) + 1 \right] \\
& + \beta_5 \cdot \ln[R\&D_{i,t-1} + 1] + \beta_6 \cdot \ln[VC_{i,t-1} + 1] + Controls + \varepsilon \quad (1)
\end{aligned}$$

where the outcome variable is either the patent count, the number of cites, the patent originality or generality of a company. We apply the inverse hyperbolic sine transformation to the outcome variables but the results are quantitatively similar with a  $\ln+1$  transformation. The second and third term is the distance weighted sum of R&D spending or VC investment the spillover pool, respectively. We use as distance metric the citation-augmented distance measure  $Tech_{i,j}^{Cites}$  defined above for the exposition but show results for all types of distance metrics. A major problem we have to address is that the size of R&D and VC investment are the outcome of an optimization process of the company and therefore endogenous. There are two sources of endogeneity which we address in turn: unobserved heterogeneity of the company and unobserved technological progress over time or technology classes.

Companies might differ in their unobserved research productivity because they have better or worse researchers or some companies might have already a patent on a particular basis technology and thus might be the only one to exploit the cumulative nature of innovations. By the cumulative nature of the innovative process, past shocks might influence current expenditures for innovation and therefore variables are only pre-determined, not strictly exogenous. Thus we resort to pre-sample mean scaling to account for firm-fixed effects. We also include in our estimation a complete set of year and industry dummies.

Unobserved technological progress might drive investment in innovation and patenting activity. This can cause a spurious correlation between our output measures and the spillover or investment variables. To alleviate this problem we use a two-step instrumental variable strategy: First we predict R&D spending and venture capital funding with the help of factors which are plausible to be exogenous to technological progress. With these predicted values we calculate the spillover terms and then use the predicted values as instruments in the equation (1)

To isolate exogenous variation in R&D expenditures we use local supply side shock caused by the staggered introduction of R&D tax-credits across states in the US follow-

ing Bloom, Schankerman, and Van Reenen (2013). These tax-credits lower the cost to do R&D and therefore should in equilibrium increase its optimal level. The literature surveyed in Bloom, Schankerman, and Van Reenen (2013) suggests that there is a degree of randomness in the introduction and the level of R&D tax-credits across states and therefore it is plausible that a change in the instrument is exogenous to technological progress.

As instrument for venture capital spending we use fund-raising of leveraged buyout funds one year before the investment following Nanda and Rhodes-Kropf (2012). The supply of venture capital is greatly influenced by the asset allocation of institutional investors into “private equity”, the broad category encompassing venture capital and buyout funds. By using buyout fund-raising we capture that part of VC investments which is due to increases in available capital unrelated to technological opportunities of venture capital-backed start-ups.

## **4. Results**

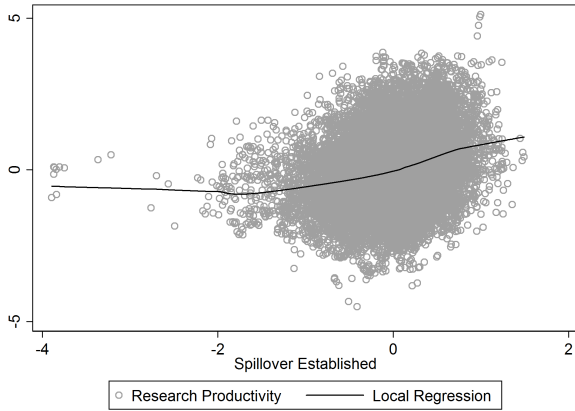
### **4.1. Descriptive Evidence**

In our framework, spillovers change the productivity of research of a company such that, for a given sum of R&D investment, a company produces more and/or better patents if it receives more spillovers. In this section we take a first look at the data to see whether this is indeed the case. We first calculate how many patents are to be expected given the investment in R&D expenditures, the industry and the year. We then look at the residual, i.e. the deviation from the true number of patents from the expected number of patents. If it is indeed true that spillovers improve the research productivity, then there should be a positive association between the spillover measures and this residual. In Figure 6 we plot the patent productivity residuals for several patent measures and the spillovers measures which we also correct for R&D expenditure, industry and year. We find a clear positive association between these residuals and spillovers from both established companies (left side of the panel) and from venture capital financed companies (right side of the panel).

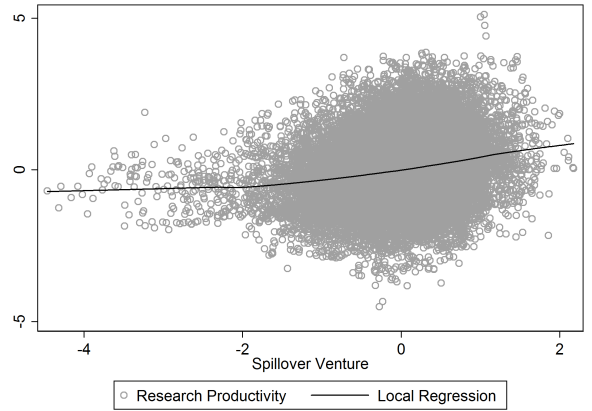
### **4.2. Summary of the Main Results**

Table 2 reports our main results for the citation-augmented distance measure. We use three different outcome measures: the total number of patents in a given year, the number of scaled forward cites and the scaled generality of patents. We use for each dependent

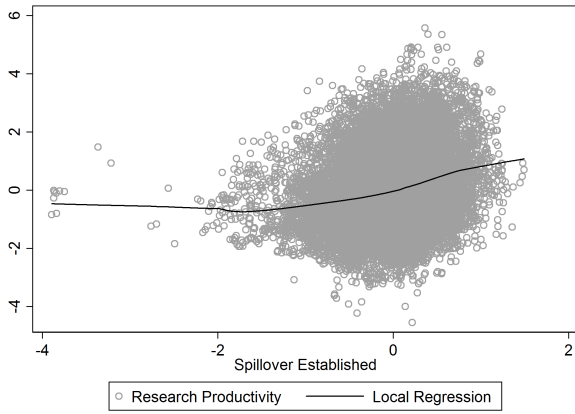
Figure 6: Research Productivity and Spillover



(a) Patent Count



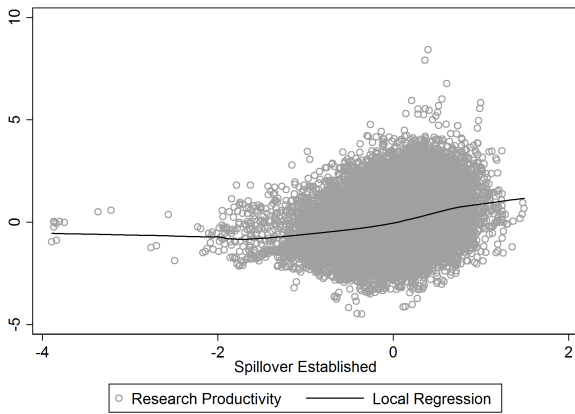
(b) Patent Count



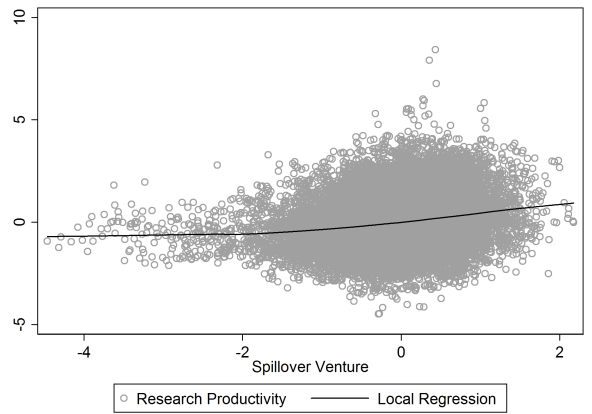
(c) Forward Cites



(d) Forward Cites



(e) Generality



(f) Generality

variable three different estimation methods: In the first column we report OLS results, then we add firm fixed effects and in the third column we use IV estimation to account for endogeneity.

All coefficients for the OLS and the firm-fixed effect regression are significantly different from zero on conventional level. This provides prima facie evidence for technology driven spillovers, but — as discussed above — the coefficients might just take up the influence of technological progress. If this is the case our estimated coefficients overstate the true effect. To account for this potential endogeneity, we report in the third column for each outcome variable an estimation with instrumental variables. Except the coefficient for own venture capital stock all coefficients are significantly different from zero on conventional levels and have a similar size across specifications.

As both the dependent and the independent variable are in logs, the estimates should be interpreted as a change in percentage terms. Therefore, according to the IV specification, an increase in the stock of R&D of 10% results in an increase in the number of patents of 3.21%. The same percent increase in venture capital yields 1.94% more patents. In percentage terms spillovers are much smaller for venture capital financed companies than for established companies: a 10% increase in spillovers of established companies yields 2.03% more patents whereas a 10% increase in spillovers of venture capital financed companies results in 1.05% more patents. Yet, the spillover measure of venture capital has with 45.14 a much smaller average value than the spillover measure for established companies with 6742.07. Consequently a ten percent increase is in dollar terms much larger for spillovers originating from R&D than for spillover originating from venture capital.

To gauge the relative sizes of the different effects we conduct back of envelope calculations for two different counterfactual scenarios and report the results in Panel A of Table 3. In the first scenario we increase the amount of R&D or venture capital investment for one year for one company and calculate given the estimated coefficients the resulting increase in patents for all other companies. We repeat this step for all companies and then take the average over all resulting change in patents. In the second scenario we do the same, except that we increase investment by 10%. Note that an average venture capital financed company has on average 1 million US \$ per year VC investment while on average an established companies has a yearly R&D budget of 42 Million US \$. Thus the first scenario constitutes in percentage terms a large increase in investment for the VC-financed company and a small increase for the established company. In the second scenario the considered dollar amounts are quite unequal for the different types of companies. For further details of this calculation please refer to the Appendix A.

A one million US dollar increase in R&D investment yields around 0.07 more patents in an established company while it results in around 0.36 additional patents for a venture capital financed company. In other companies (measured according to our citation-augmented distance measure), the increase is 0.01 for an established company and 0.42 for a venture capital financed company. On average 14% of all patent production is an external effect if an established company invests while it is more than 54% if a venture capital financed company invests. In the second scenario the numbers for established companies are much higher (and so is investment) and the share of external patent production increases to 43% for established companies.

In these calculations, the estimated impact of venture capital might be too large because we only consider companies which patented at least once. Thus we neglect all investment of companies which did not result in at least one patent. Yet, we also do not know if these companies were trying to patent at all. To ameliorate this problem we calculate a correction factor by industry and year. The correction factor is the total sum of VC investment in our sample divided by the total sum VC investment in the database. Then we divide the total investment by this factor. Thus we artificially increase investment in industries which are unsuccessful in patenting, arriving at a lower bound for the effect of VC spillovers. We recalculate all the coefficients of the main result tables and then use these coefficients to calculate for the two counterfactual scenarios. We report the results in Panel B of Table 3.

The spillover effects are much smaller in this specification but still quite large: An increase in million US\$ venture capital investment results in 0.13 patents for other companies while the same increase in R&D results only 0.01 patents. The social return of venture capital investment is around 7.75 times the social return of R&D investment. This estimate is larger than the estimates found in the literature: Kortum and Lerner (2000) and Popov and Roosenboom (2012) find both that the social return of venture capital investment is around three times larger in terms of patents than the social return of R&D.

### **4.3. Robustness: Alternative definitions of the spillover pool and different outcome variables**

As discussed in section 3, there are two alternative ways to construct the spillover pool by using the Jaffe and the Mahalanobis distance measure and we results using these measures are reported in Table 4. The only difference as compared to the results of the citation-augmented distance measure is that the coefficient for venture capital spillovers is markedly smaller and not significantly different from zero on conventional levels. All

Table 2: Coefficient estimates using the citation-augmented distance

	<i>Patent Count</i>			<i>Cites</i>			<i>Generality</i>		
	OLS	FE	IV	OLS	FE	IV	OLS	FE	IV
ln(Spillover Est.)	27.8*** (3.08)	21.0*** (2.69)	20.3*** (2.81)	27.2*** (2.95)	20.9*** (2.63)	21.0*** (2.78)	30.2*** (3.44)	22.9*** (3.02)	22.8*** (3.20)
ln(Spillover VC)	8.6*** (2.05)	8.1*** (1.88)	10.5*** (3.10)	7.9*** (2.21)	7.2*** (2.08)	9.4*** (3.54)	8.9*** (2.45)	8.7*** (2.30)	9.7*** (3.55)
ln(R&D Stock)	48.4*** (1.38)	39.4*** (1.58)	32.1*** (5.77)	47.3*** (1.46)	37.1*** (1.54)	24.8*** (6.64)	49.0*** (1.48)	39.7*** (1.57)	31.4*** (5.97)
ln(VC Stock)	13.1*** (1.26)	9.2*** (1.32)	19.4 (12.92)	14.7*** (1.52)	11.3*** (1.56)	33.7** (13.77)	13.5*** (1.60)	9.5*** (1.65)	41.0*** (13.63)
ln(Sale+1)	12.2*** (1.24)	7.9*** (1.19)	13.2*** (4.20)	11.7*** (1.35)	8.1*** (1.27)	17.1*** (4.86)	11.3*** (1.36)	7.6*** (1.31)	14.2*** (4.46)
Pre-sample FE		40.7*** (1.92)	42.1*** (2.28)		42.4*** (1.66)	44.5*** (2.13)		41.2*** (1.81)	42.2*** (2.12)
F-Value	.	.	28.48	.	.	28.45	.	.	28.54
R2	0.53	0.57	0.57	0.46	0.51	0.50	0.47	0.50	0.49
N	22071	22071	22071	22071	22071	22071	22071	22071	22071

**Note:** All standard errors are clustered on the industry-year level. \*\*\*, \*\*, \* indicate statistical significance at the 1%, 5%, and 10% level, respectively. To increase the readability of the table we multiply each estimate by 100.

Table 3: Counterfactual increase in patents

**Panel A: As-is VC spending**

*Scenario 1: Increase spending by 1 Million Dollar*

	External Return	Private Return	Social Return	Share
Established	0.01	0.07	0.08	0.14
VC	0.42	0.36	0.78	0.54

*Scenario 2: Increase spending by 10% of current investment*

	External Return	Private Return	Social Return	Share
Established	0.28	0.37	0.65	0.43
VC	0.21	0.09	0.30	0.69

**Panel B: Adjusted VC spending**

*Scenario 1: Increase spending by 1 Million Dollar*

	External Return	Private Return	Social Return	Share
Established	0.01	0.03	0.04	0.29
VC	0.13	0.30	0.43	0.31

*Scenario 2: Increase spending by 10% of current investment*

	External Return	Private Return	Social Return	Share
Established	0.33	0.17	0.50	0.65
VC	0.20	0.05	0.25	0.78



other coefficient estimates are very similar in size and significance as compared to the results reported in Table 2. Apparently, including the information on cross citations to define the spillover pool improves the measureability of spillovers for venture capital financed companies.

In Table 5 we report other outcome variables. The first column reports the estimates for the originality of the patents of a company. The originality of a patent is one minus the Herfindahl-index over the shares in different patent classes from which the patents cites. In column two and three, we use the probability that a patent has more cites than the median and the 99% quintile in one year and patent class as outcome variable. In the last three column we report the unscaled version of all the outcome variables used in the main specification. All resulting estimates are consistent with our results in the main specification.

#### **4.4. What makes venture capital-financed companies different?**

We measure higher spillovers from VC financed start-ups than from regularly financed R&D. There might be two possible reasons for this: The VC investment process itself might cause the start-ups to exhibit large spillovers, a treatment effect. This means VC fund manager have a way to improve the research of their investee companies. Second, venture capitalist might choose to invest in start-ups which have high spillovers ex-ante, a selection effect. For example VC backed start-ups might commercialize new technologies with large growth potential, but also a lot of risk.

In this section we present tentative evidence that VC backed start-ups are already very innovative ex-ante and central in the technology space. This points to an selections effect, as venture capitalist only can influence a company after investment and at the time of first investment the general technology position is already fixed. This is consistent with the findings in the literature that venture capital is invested in innovative firms (Caselli, Gatti, and Perrini, 2009; Hellmann and Puri, 2000; Engel and Keilbach, 2007). Yet, the IV results in table 2 imply that an exogenous increase in VC financing leads to more spillover, establishing a causal effect of venture capital financing. This might imply a selection effect on the extensive margin: In the absence of VC funding these companies would not have received funding at all and no patents would have existed.

##### **4.4.1. VC financed companies are more central**

In table 6 we divide all companies in quantiles according to the total number of patents they have in our sample. Then we take the average of the distance to all other companies

Table 4: Coefficient estimates using alternative definitions of the spillover pool

	Jaffe Distance								
	<i>Patent Count</i>			<i>Cites</i>			<i>Generality</i>		
	OLS	FE	IV	OLS	FE	IV	OLS	FE	IV
ln(Spillover Est. (Jaffe))	32.8*** (2.25)	24.6*** (2.24)	26.8*** (2.68)	31.1*** (2.32)	23.4*** (2.29)	27.4*** (2.81)	36.0*** (2.62)	27.2*** (2.57)	31.2*** (3.17)
ln(Spillover VC (Jaffe))	-0.4 (1.49)	0.8 (1.42)	1.3 (2.57)	-0.4 (1.58)	0.8 (1.53)	-0.1 (2.79)	-1.2 (1.78)	0.4 (1.72)	-1.2 (2.89)
ln(R&D Stock)	49.1*** (1.37)	40.0*** (1.57)	28.1*** (6.61)	48.0*** (1.44)	37.7*** (1.52)	22.8*** (7.18)	49.8*** (1.46)	40.5*** (1.55)	28.5*** (6.57)
ln(VC Stock)	13.4*** (1.23)	9.4*** (1.29)	-1.1 (15.39)	15.0*** (1.49)	11.5*** (1.54)	16.1 (16.51)	13.9*** (1.57)	9.8*** (1.62)	20.9 (16.39)
ln(Sale+1)	11.8*** (1.28)	7.6*** (1.23)	15.6*** (4.73)	11.4*** (1.38)	7.8*** (1.31)	18.3*** (5.26)	10.9*** (1.40)	7.2*** (1.35)	15.9*** (4.85)
Pre-sample FE		40.6*** (1.92)	43.6*** (2.49)		42.3*** (1.67)	45.2*** (2.33)		41.0*** (1.80)	43.3*** (2.34)
F-Value	.	.	25.01	.	.	25.11	.	.	25.06
R2	0.53	0.57	0.57	0.46	0.51	0.51	0.46	0.50	0.50
N	22071	22071	22071	22071	22071	22071	22071	22071	22071

**Note:** All standard errors are clustered on the industry-year level. \*\*\*, \*\*, \* indicate statistical significance at the 1%, 5%, and 10% level, respectively. To increase the readability of the table we multiply each estimate by 100.

	Mahalanobis Distance								
	<i>Patent Count</i>			<i>Cites</i>			<i>Generality</i>		
	OLS	FE	IV	OLS	FE	IV	OLS	FE	IV
ln(Spillover Est. (Mal))	45.7*** (3.46)	33.7*** (3.54)	36.1*** (4.09)	43.5*** (3.50)	32.3*** (3.56)	35.6*** (4.16)	51.1*** (3.93)	38.0*** (3.96)	41.9*** (4.69)
ln(Spillover VC (Mal))	-1.1 (2.39)	1.3 (2.34)	0.5 (3.63)	-0.9 (2.50)	1.3 (2.47)	-0.3 (3.99)	-2.5 (2.81)	0.5 (2.75)	-2.3 (4.16)
ln(R&D Stock)	48.5*** (1.35)	39.7*** (1.56)	34.6*** (5.91)	47.4*** (1.41)	37.3*** (1.52)	28.7*** (6.58)	49.1*** (1.43)	40.1*** (1.54)	34.0*** (5.90)
ln(VC Stock)	12.7*** (1.20)	9.0*** (1.26)	5.6 (12.89)	14.4*** (1.48)	11.1*** (1.53)	22.2 (13.89)	13.2*** (1.55)	9.3*** (1.60)	26.8** (13.55)
ln(Sale+1)	10.7*** (1.26)	6.9*** (1.21)	10.3** (4.29)	10.3*** (1.36)	7.1*** (1.29)	13.3*** (4.87)	9.6*** (1.37)	6.5*** (1.33)	11.1** (4.43)
Pre-sample FE		39.7*** (1.94)	40.9*** (2.38)		41.7*** (1.70)	43.3*** (2.21)		40.0*** (1.81)	40.9*** (2.20)
F-Value	.	.	29.47	.	.	29.34	.	.	29.63
R2	0.54	0.57	0.57	0.47	0.51	0.51	0.47	0.50	0.50
N	22071	22071	22071	22071	22071	22071	22071	22071	22071

**Note:** All standard errors are clustered on the industry-year level. \*\*\*, \*\*, \* indicate statistical significance at the 1%, 5%, and 10% level, respectively. To increase the readability of the table we multiply each estimate by 100.

Table 5: Other outcomes

	Other Scaled Outcomes			Unscaled		
	<i>Originality</i>	<i>Forward Cites &gt;median</i>	<i>Forward Cites &gt;99%</i>	<i>Forward Cites</i>	<i>Gen.</i>	<i>Orig.</i>
ln(Spillover Est.)	23.5*** (3.21)	19.1*** (2.59)	2.6*** (0.57)	36.7*** (4.97)	11.8*** (1.79)	15.1*** (2.28)
ln(Spillover VC)	10.2*** (3.62)	10.3*** (3.15)	2.9*** (0.76)	16.7** (6.50)	9.5*** (2.35)	10.1*** (2.53)
ln(R&D Stock)	35.0*** (6.29)	32.2*** (5.64)	0.3 (1.56)	67.8*** (11.39)	13.0*** (3.89)	21.6*** (4.65)
ln(VC Stock)	25.9* (14.38)	24.2** (11.05)	-8.7*** (2.79)	123.8*** (22.37)	40.2*** (10.99)	-0.9 (10.53)
Pre-sample FE	41.4*** (2.21)	2.6*** (0.18)	23.7*** (1.79)	0.1*** (0.01)	36.2*** (2.02)	37.5*** (1.90)
F-Value	28.17	27.87	29.22	28.30	28.54	28.17
R2	0.54	0.53	0.28	0.38	0.53	0.58
N	22071	22071	22071	22071	22071	22071

for each company and then take the median within each quantile. Companies with more patents are in the spillover pool of more other companies because their position is technologically more diverse. Conditioned on the number of patents venture capital backed start-ups are much more central than established companies: For example an established company with 4 patents has a proximity of 1.75 to all other companies while the proximity of a similar sized VC financed company is 2.52. Only the largest established companies are closer to other companies than the largest VC financed companies, however the latter have only 5% of the established companies patents. This central position can only be rationalized with treatment effect if the investor would be able to change the technological position of the investee company, i.e. the patent classes in which a company patents. Kaplan, Sensoy, and Stroemberg (2009) points out that business plans of successful companies are stable over time, what might imply that the position in technology space is so too.

#### 4.4.2. VC financed companies do better research (but it becomes worse over time)

In table 7 and figure ?? we show the characteristics of patents of established companies and venture backed start-ups. The patents of VC backed start-ups we split into a sample of innovation before and after the first investment. Venture capital backed start-ups are receive more cites, cite more other patents, are more general and more original than patents of established companies. Compared to established companies, VC companies

Table 6: Median Distance

	Quintile	Patent Count	Employment (Thsd)	Proximity (Jaffe)	Proximity (Cites)
Established	1	4	0.26	1.74	1.75
Established	2	15	0.47	2.17	2.19
Established	3	54	1.04	2.45	2.45
Established	4	440	6.99	3.06	3.34
Venture Capital	1	1	0.00	2.18	1.93
Venture Capital	2	4	0.00	2.57	2.52
Venture Capital	3	9	0.00	3.22	2.77
Venture Capital	4	22	0.00	3.51	3.12

patent in areas which are more active (measured by the number of other patents applied for in this class). Also patents of venture capital financed start-ups are more likely to be in the upper tail of the distribution of each patent characteristics according to Table 8. In this table we calculate the probability that the patents of established and venture capital financed start-ups are at or above the 99% quintile in this characteristic, i.e. if they are outlier. If all patents of each group would be equally likely to be an outlier, the average propensity should be one. However, for VC financed companies it is above 1.9% while it is below 1% for established companies. So it appears that research done in venture capital backed start-ups is more innovative and used in diverse technology fields from a lot of other patents.

After the first investment the patent of start-ups are on average less cited and less general while the number patents does not decrease (Figure 7 and Figure 8). This age effect is stable after controlling for application year and patent class as shown in Table 9. The high quality research before the first investment cannot be due to a treatment effect of the venture capital investor but can only come from a selection effect. Yet, this does not establish that there is no treatment effect at all as we do not observe the research performance of companies after the first investment if there would have not been an venture capital investor involved.

Taking these and the IV results together suggests, that there might be a selection effect on the extensive margin: VC investors choose companies with high spillover, but if they are not selected these companies cannot receive alternative funding. Yet the interpretation applies strictly only for start-ups which are shifted in and out by the additional funding measured by the instrumental variable. Companies which receive funding at all times might be of such high quality, that they can get investment from alternative sources. If this is the case, the spillover effects might be a pure selection effect.

Table 7: Mean Characteristics of Venture Capital and Established Patents

	Established	VC (before investment)	VC (after)
Forward Cites	11.38	19.90	15.43
Scaled Forward Cites	0.931	1.585	1.296
Generality	0.288	0.365	0.297
Scaled Generality	0.216	0.278	0.227
Originality	0.399	0.420	0.434
Scaled Originality	0.287	0.305	0.313
Backward Cites	14.20	21.66	23.57
Scaled Backward Cites	1.069	1.413	1.509
Other Patents in the same class/year	1273.5	1473.4	1717.2
Total Number	686057	2978	28596

Figure 7: Patent characteristics of start-up after receiving the first investment

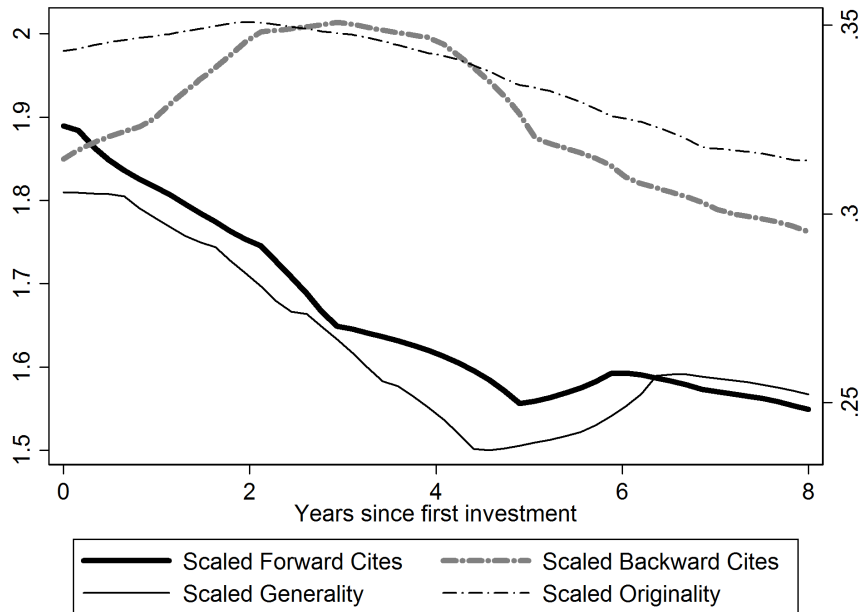


Figure 8: Average number of patents per start-up and years since investment

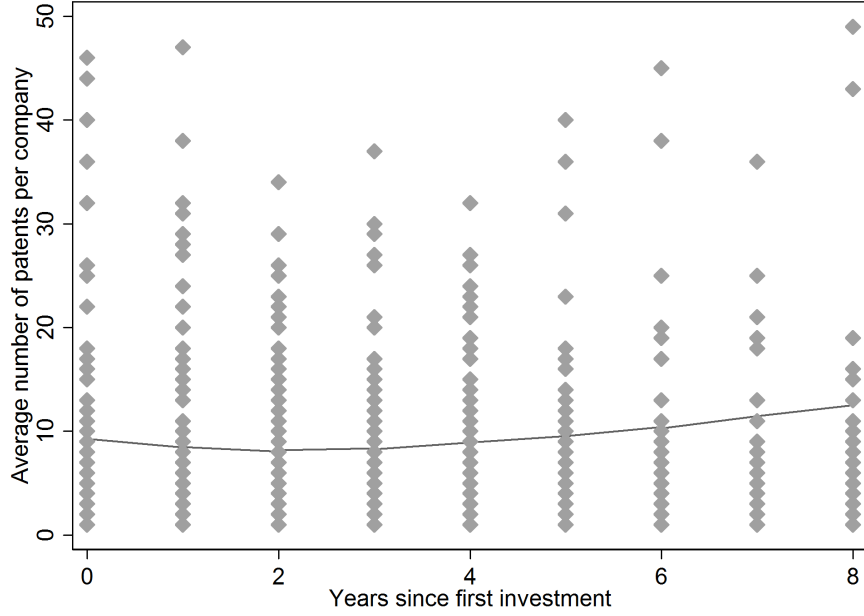


Table 8: Probability of a patent to be in 99% quantile of the distribution

	Established	VC (before investment)	VC (after investment)
Forward Cites	0.934	3.156	2.432
Scaled Forward Cites	0.932	3.526	3.017
Generality	0.983	2.418	1.397
Scaled Generality	0.951	2.955	2.448
Originality	0.948	1.948	2.629
Scaled Originality	0.945	1.948	2.810
Backward Cites	0.903	2.921	3.913
Scaled Backward Cites	0.924	2.317	3.488
Total Number	686057	2978	19322

Table 9: Correlation of start-up age and scaled patent characteristics

	Forward Cites	Generality	Backward Cites	Originality
Years since first investment	-2.50** (1.07)	-0.19** (0.08)	0.19 (1.23)	-0.18** (0.08)
Patent Class FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R2	0.11	0.24	0.10	0.14
N	14510	14510	14510	14510

## 5. Conclusion

In this paper we have shown that VC-financed firms generate significant and positive spillovers on other firms' patent quantity and quality. Moreover, counterfactual simulations suggest that the relative share of external versus private returns is larger for VC-financed firms than for established companies. We also find that the Jaffe or Mahalanobis distance measures fail to capture relevant knowledge flows of VC-financed start-ups. In comparison, a citation-augmented distance measure is better able to capture the spillover effects of VC-financed firms as identified by the backward citation pattern.

## References

- BERNSTEIN, S. (2012): “Does Going Public Affect Innovation?,” Mimeo.
- BLOOM, N., M. SCHANKERMAN, AND J. VAN REENEN (2013): “Identifying Technology Spillovers and Product Market Rivalry,” *Econometrica*, 81(4), 1347–1393.
- CASELLI, S., S. GATTI, AND F. PERRINI (2009): “Are venture capitalists a catalyst for innovation?,” *European Financial Management*, 15(1), 92–111.
- DENG, Y. (2008): “The value of knowledge spillovers in the US semiconductor industry,” *International Journal of Industrial Organization*, 26(4), 1044–1058.
- ENGEL, D., AND M. KEILBACH (2007): “Firm-level implications of early stage venture capital investment—An empirical investigation,” *Journal of Empirical Finance*, 14(2), 150–167.
- GRIFFITH, R., S. LEE, AND J. VAN REENEN (2011): “Is distance dying at last? Falling home bias in fixed-effects models of patent citations,” *Quantitative Economics*, 2(2), 211–249.
- HALL, B. H., A. JAFFE, AND M. TRAJTENBERG (2005): “Market value and patent citations,” *RAND Journal of economics*, 1, 16–38.
- HALL, B. H., A. B. JAFFE, AND M. TRAJTENBERG (2001): “The NBER patent citation data file: Lessons, insights and methodological tools,” Discussion paper, National Bureau of Economic Research.
- HALL, B. H., J. MAIRESSE, AND P. MOHNEN (2009): “Measuring the Returns to R&D,” Working Paper 15622, National Bureau of Economic Research.
- HELLMANN, T., AND M. PURI (2000): “The interaction between product market and financing strategy: the role of venture capital,” *Review of Financial Studies*, 13(4), 959–984.
- JAFFE, A. B. (1986): “Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits, and Market Value,” *The American Economic Review*, 76(5), 984–1001.
- JAFFE, A. B., M. TRAJTENBERG, AND M. S. FOGARTY (2000): “Knowledge spillovers and patent citations: Evidence from a survey of inventors,” *NBER/Sloan*, p. 21.



- JAFFE, A. B., M. TRAJTENBERG, AND R. HENDERSON (1993): “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *The Quarterly Journal of Economics*, 108(3), 577–598.
- KAPLAN, S. N., B. A. SENSOY, AND P. STROEMBERG (2009): “Should Investors Bet on the Jockey or the Horse? Evidence from the Evolution of Firms from Early Business Plans to Public Companies,” *The Journal of Finance*, 64(1), 75–115.
- KORTUM, S., AND J. LERNER (2000): “Assessing the Contribution of Venture Capital to Innovation,” *The RAND Journal of Economics*, 31(4), pp. 674–692.
- LAI, R., A. AMOUR, A. YU, Y. SUN, V. TORVIK, AND L. FLEMING (2011): “Disambiguation and co-authorship networks of the US Patent Inventor Database,” *Harvard Institute for Quantitative Social Science, Cambridge, MA*, 2138.
- NANDA, R., AND M. RHODES-KROPF (2012): “Investment Cycles and Startup Innovation,” forthcoming, *Journal of Financial Economics*.
- POPOV, A., AND P. ROOSENBOOM (2012): “Venture capital and patented innovation: evidence from Europe,” *Economic Policy*, 27(71), 447–482.
- TRAJTENBERG, M., R. HENDERSON, AND A. JAFFE (1997): “University versus corporate patents: A window on the basicness of invention,” *Economics of Innovation and new technology*, 5(1), 19–50.

## A. Back of the envelope estimation: Details

Calculating the economics significance of the spillover effects on the basis of our estimates is complex for three reasons: First, conceptually, without specifying a concrete policy to consider it is unclear what constitutes an interesting counterfactual. For example, an increase in investment of one million dollar is extremely large for a start-up company while it is small for a larger firm whose R&D spending is already over half a billion. For this reason we take a two-pronged approach: We calculate one scenario where we increase the investment in innovation of one company by one million dollar and then take the average over the resulting patents for all companies. This is a comparably large increase for venture capital financed companies given that the average company has 1 Million US \$ spending per year but a rather small increase for an established companies which has on average a yearly R&D budget of 42 Million US \$. In the second scenario, we increase investment by 10%, which means 4 Million US \$ for the average established company and 100'000 US \$ for an average VC-financed company.

Second, a technical problem with calculating spillover effects is that one has to take into account firm heterogeneity because spillovers vary with the size of the company and its position in technology space. For example, according to the first column in Table 2 increasing the R&D capital stock of a company by 10% increases the number of patents by 4.84 %. Obviously, 10% percent are quite different in dollar terms for a large and a small company. Furthermore, spillover originating from a company depend on its position in technology space. A company which is technologically closer to other companies has more spillovers under any counterfactual scenario. For this reason we re-calculate for both scenarios (one million dollar or 10% more investment) the change of all counterfactual variables for one company at a time to predict the increase in patent production.

Third, the estimated impact of venture capital might be too large because we only consider companies which patented at least once. Thus we neglect all investment of companies which did not result in at least one patent. Yet, we also do not know if these companies were trying to patent at all. To ameliorate this problem we calculate a correction factor by industry and year. The correction factor is the total sum of VC investment in our sample divided by the total sum VC investment in the database. Then we divide the total investment by this factor. Thus we artificially increase investment in industries which are unsuccessful in patenting, arriving at a lower bound for the effect of VC spillovers. We re-calculate all the coefficients (Table 8) of the main result tables and then use these coefficients to calculate for the two counterfactual scenarios.

Table 10: Coefficient estimates for adjusted VC spending

Outcome Method	<i>Patent Count</i>			<i>Cites</i>			<i>Generality</i>		
	OLS	FE	IV	OLS	FE	IV	OLS	FE	IV
ln(Spillover Est.)	30.7*** (3.31)	23.3*** (2.91)	21.7*** (3.08)	29.8*** (3.19)	23.0*** (2.86)	22.5*** (3.07)	33.3*** (3.68)	25.2*** (3.26)	24.5*** (3.51)
ln(Spillover VC Corr)	4.5** (2.09)	4.6** (1.92)	7.7** (3.14)	4.0* (2.25)	4.2** (2.10)	6.9* (3.54)	4.4* (2.48)	5.1** (2.31)	7.1** (3.60)
ln(R&D Stock)	49.0*** (1.38)	39.8*** (1.58)	32.5*** (5.81)	47.8*** (1.46)	37.4*** (1.56)	24.3*** (6.67)	49.6*** (1.48)	40.2*** (1.59)	30.5*** (6.09)
ln(VC Stock)	13.3*** (1.25)	9.3*** (1.31)	25.8** (11.40)	14.9*** (1.50)	11.5*** (1.55)	37.8*** (11.90)	13.7*** (1.59)	9.7*** (1.64)	44.6*** (11.65)
ln(Sale+1)	11.9*** (1.25)	7.7*** (1.20)	13.1*** (4.24)	11.5*** (1.36)	7.9*** (1.28)	17.6*** (4.89)	11.1*** (1.37)	7.5*** (1.32)	15.0*** (4.54)
Pre-sample FE		40.8*** (1.92)	41.9*** (2.28)		42.4*** (1.67)	44.7*** (2.10)		41.2*** (1.81)	42.5*** (2.07)
F-Value	.	.	40.15	.	.	39.88	.	.	40.38
R2	0.53	0.57	0.57	0.46	0.51	0.50	0.47	0.50	0.49
N	22071	22071	22071	22071	22071	22071	22071	22071	22071