Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

# The Location of Academic Institutions and Knowledge Flow to Industry: Evidence from Simultaneous Discoveries

**michael Bikard**
**lbs**
**entrepreneurship**
**bikard@london.edu**

**Matt Marx**
**MIT**
**Sloan School of Management**
**mmarx@mit.edu**

## Abstract

Scientific discoveries in academia can spur innovation and growth, but only if they flow to industry. This paper documents a source of friction to the flow of academic science to firms: corporate inventors tend to systematically overlook academic discoveries that emerge outside concentrations or "hubs" of relevant industrial R&D. Testing the impact of location on knowledge flow is difficult because institutions at different locations produce different kinds of research. We address this problem by analyzing simultaneous discoveries where multiple researchers report the same finding in separate "twin" papers. Even after accounting for the localization of spillovers, we find that "twin" papers conducted outside of a hub of relevant R&D are approximately 10% less likely to be referenced as prior art by firm-assigned patents. This effect is moderated by collocation, the institution's prestige as well as formal connections with industry. Taken together, our results suggest that the geographic location of academic institutions affects the chances that their discoveries become orphaned, with sobering implications for the science of science policy but also potential opportunities for firms commercializing academic science.

Jelcodes:O31,O33

# The Location of Academic Institutions and Knowledge Flow to Industry: Evidence from Simultaneous Discoveries

**Abstract:** Scientific discoveries in academia can spur innovation and growth, but only if they flow to industry. This paper documents a source of friction to the flow of academic science to firms: corporate inventors tend to systematically overlook academic discoveries that emerge outside concentrations or "hubs" of relevant industrial R&D. Testing the impact of location on knowledge flow is difficult because institutions at different locations produce different kinds of research. We address this problem by analyzing simultaneous discoveries where multiple researchers report the same finding in separate "twin" papers. Even after accounting for the localization of spillovers, we find that "twin" papers conducted outside of a hub of relevant R&D are approximately 10% less likely to be referenced as prior art by firm-assigned patents. This effect is moderated by collocation, the institution's prestige as well as formal connections with industry. Taken together, our results suggest that the geographic location of academic institutions affects the chances that their discoveries become orphaned, with sobering implications for the science of science policy but also potential opportunities for firms commercializing academic science.

Academic research is an essential engine of innovation and growth (Romer 1990; Grossman and Helpman 1993; Aghion, Dewatripont, and Stein 2008), with governments across the world investing billions annually with the expectation that economic benefits will follow. Although academic research can be used by firms to increase R&D efficiency (Nelson 1959; Nelson 1982; Cohen, Nelson, and Walsh 2002; Mokyr 2002), these benefits accrue only if this knowledge flows to industry. In the early 19$^{th}$ century, Charles Babbage highlighted the crucial role of science for "the arts and manufactures" and argued that the connection between science and those manufactures "should be rendered more intimate" (Babbage 1832, 307). Such concerns are not just theoretical: NIH director Francis Collins recently declared that he was "frustrated to see how many of the [academic] discoveries that do look as though they have therapeutic implications are waiting for the pharmaceutical industry to follow through with them" (Harris 2011). This observation clearly reflects the need for greater understanding of the circumstances under which valuable academic knowledge fails to flow to industry.

The location of academic research institutions might be a source of frictions in the flow of knowledge to industry because their geographic distribution rarely matches that of firms. Thus academic scientists are generally located far from the firms that could use their research results. The localization of knowledge spillovers (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Adams 2002; Furman and MacGarvie 2007; Azoulay, Graff Zivin, and Sampat 2012; Belenzon and Schankerman 2013) suggests that this distance could generate important inefficiencies. Distance, however, might not affect every academic institution equally. Consider a San Diego-based biotech firm building on scientific knowledge discovered simultaneously by academic researchers in both Dallas, Texas and Boston, Massachusetts. One might expect that the biotech firm will be more likely to reference the work from Dallas rather than Boston since it is closer. However, we predict that the firm will exploit the Boston-based discovery, even though it is further away, because Boston is more a "hub" of biotech R&D than is Dallas. Academic research conducted inside relevant R&D hubs is likely to receive systematically more attention than research conducted outside such a hub. Hence, we propose that the location of an academic institution inside a hub can offset distance as a constraint on diffusion; by the same token, valuable knowledge outside of such hubs may be ignored.

Our hypothesis is not straightforward to examine empirically. Industry inventors might ignore scientific discoveries located outside of R&D hubs, but this might instead occur because such discoveries are less applied or perhaps more "fundamental"—i.e., not immediately useful to firms. Although we are interested in the marginal impact of location of the academic institution on knowledge flow to industry, a selection effect may confound our analysis if academic knowledge that is technologically more useful emerges in institutions that are located in R&D hubs. Certainly, recent evidence suggests that the direction of academic science is in part endogenous to local firms' research priorities (Sohn 2014).

To address these issues, we exploit the occurrence of simultaneous scientific discoveries (Merton 1961). When two or more scholars publish their findings at about the same time, they create "paper twins." We measure the flow of academic discoveries to industry by observing references from firm-assigned patents to paper twins disclosing simultaneous discoveries. Our identification strategy is based on differential referencing of one paper twin versus another, where one twin is within a hub of relevant R&D and the other is not. Our empirical strategy has three key advantages.

First, the use of patent references as a measure of knowledge flow is complicated by the possible existence of false positives as citations are often added ex-post for legal or strategic reasons (Alcácer, Gittelman, and Sampat 2009; Lampe 2012). Our identification strategy is facilitated by a key feature of the USPTO rules for recognizing prior art. Although inventors are required to disclose all relevant prior art, Rule 56 states that an inventor is not required to reference multiple sources disclosing the same prior art. Thus if a simultaneous discovery were relevant prior art, but the patent referenced only one of the "twin" papers reporting that discovery, not referencing another twin paper would not affect either the scope or the validity of the patent.

Second, since academic discoveries tend to be published and not patented, a focus on references to academic patents might lead to a considerable number of false negatives. We circumvent this difficulty by using patent references to scientific articles. While scientific references from patents by no means capture every flow of academic knowledge to industrial R&D, Roach and Cohen (2013) report that they are perhaps the most reliable indicator.

Third, as in any case-control analysis the reliability of inference depends critically on the quality of the controls. Much work on knowledge diffusion uses coarse controls that could confound inference (Thompson and Fox-Kean 2005), so our ability to observe the same discovery in different contexts provides a rare opportunity to control for the quality of the underlying discovery.

We find that the location of academic research impacts its flow to industry. First, using patent references to simultaneous discoveries, we replicate the Belenzon and Schankerman (2013) finding that the flow of academic knowledge to industry is attenuated by spatial separation between the academic scientist and the firm. Second, we show that even when controlling for knowledge localization, academic discoveries made outside of industrial R&D hubs are less likely to flow to firms. This effect is not driven by distance from the focal patent: even after accounting for the distance between the paper twins and a patent that might reference them, academic discoveries that emerge outside of R&D hubs appear more likely to be overlooked by industry inventors. This result is robust to a number of different specifications and holds only for references from firm-assigned patents; references from university-assigned patents to academic papers are unaffected. Moreover, this effect is attenuated in the following circumstances: 1) institutions have formal connections to industry; 2) for papers at institutions with a higher academic reputation, which might draw attention from firms even if the institution is located outside of an R&D hub; and 3) when the focal paper and the potentially-referencing patent are themselves colocated. It therefore appears that the negative impact of being outside of an R&D hub can be mitigated when industrial inventors' attention is drawn to a paper by other means.

## 1. The Location of Academic Institutions and Knowledge Flow to Industry

### A. The Flow of Academic Science to Industry

Scientific knowledge can increase R&D efficiency because it guides the invention process. From the perspective of an inventor, R&D capability therefore depends on the person's knowledge (Nelson 1982). Joel Mokyr (2002) notes that society's understanding about nature and its regularities can help it invent because the process of invention consists in a large part in using these regularities for a purpose. In line with this reasoning, large-scale empirical studies have established a link between university research and corporate patenting (Jaffe 1989) as well as productivity growth (Adams 1990).

Despite the well-studied openness of the academic environment (Merton 1973; Dasgupta and David 1994; Stephan 1996), concerns have been voiced that frictions might prevent the flow of academic knowledge to industry. Such inefficiency might exist if those firms have imperfect access to the newly produced knowledge. Mokyr (2002) proposes that "progress in exploiting the existing stock of knowledge will depend first and foremost on the efficiency and cost of access to knowledge." This paper explores this proposition empirically—in particular, we focus on the impact of the geographic location of academic research institutions.

### B. Geographic Location and the Flow of Academic Science to Industry

Location might affect the flow of academic science to firms because academic research institutions and industrial R&D labs are not always collocated. The potential for frictions depends on the geographic distribution of the knowledge demand and supply, a topic thoroughly studied by economic geographers (for a review, see Feldman & Kogler 2010). In some locations, one might expect little friction because academic researchers might all be collocated with the relevant industry inventors, therefore facilitating knowledge flows. In most places, however, scientists are not all collocated with the firms that could use their research results, potentially leading to considerable frictions in the dissemination of knowledge.

The most studied friction in the flow of knowledge from academia to industry is probably the geographic localization of knowledge spillovers. By itself, however, the localization argument does not necessarily imply the existence of frictions as firms located further away from academic institutions might not have the same knowledge needs as those that are nearby. Therefore, to achieve a deeper understanding of the mechanism underlying the localization of academic knowledge spillovers, it is important to investigate whether the knowledge will flow as opposed to who benefits from that flow.

We examine another implication of corporate inventors' imperfect visibility into new academic knowledge: that the location of an academic research institution inside or outside of a *"hub" of relevant* industrial R&D will have important implications for knowledge flow, especially to distant firms. R&D hubs are central areas of knowledge generation and exchange among collaborators, competitors, and beyond. These are the types of locations in which, as Alfred Marshall (1895, 225) famously described "good work is rightly appreciated, inventions (…) have their merits promptly discussed: if one man starts a new idea, it is taken up by others." Far-flung industrial inventors trying to stay current the latest academic science in their field may tend to focus their attention on developments where similar commercial R&D is also happening. In contrast, academic discoveries emerging outside those hubs are unlikely to benefit from the same attention. As a result, industry inventors are presumably much more likely to overlook valuable discoveries emerging outside those hubs.

## 2. Data Construction

### A. Empirical Approach

Identifying the impact of the location of an academic research institution on knowledge flow to industry is nontrivial because the emergence of academic discoveries in specific locations is not exogenous to the geographic distribution of industrial R&D. For our purpose, this impact raises an important identification challenge. To take an example, when observing that academic research produced at MIT or at Stanford has a disproportionate impact on firms' R&D activities, how can the empiricist adjudicate whether this is because those institutions are located in R&D hubs or because the discoveries that they produce are more relevant to firms, or of higher quality?

We use simultaneous discoveries in which one team is based in a hub of corporate R&D and the other is outside such a hub. This approach is advantageous in that it still allows us to observe differences in knowledge flow from the same discovery emerging at different locations. However, since those discoveries emerge on the same planet and are published in the same set of highly visible academic journals, we should be concerned about contamination between the treatment and the control. Corporate inventors who are aware of one discovery paper are also presumably more likely to be aware of its "twin." This would introduce a bias against us finding any systematic difference in referencing, unless inventors have referencing preferences—i.e., that conditional on being aware of both twin papers, they would prefer to reference one in their patents rather than the other.

Our dataset is not a random sample of academic discoveries. The discoveries in our data are particularly important ones if only because we identified the "paper twins" by focusing on systematic co-citation in the academic literature. This means that poorly cited simultaneous discoveries would not have entered our dataset. In addition, we are focusing on cases in which inventors reference at least of the simultaneous discovery papers, which means that we are not able to observe cases in which both twin papers could have been referenced but none did. By focusing specifically on twins in which at least one but not all papers are referenced in patents, we might be selecting on a specific kind of knowledge that can flow to industry but does not flow seamlessly.

B. Illustration

Our identification strategy hinges on simultaneous scientific discoveries. Before describing the process by which these were found, we illustrate the nature of a simultaneous discovery with an example. The August 1998 issue of Cell contains two papers reporting the same scientific discovery.

| |
|---|
| **Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis**<br>Li, Zhu, Xu, and Yuan at Harvard Medical School, Boston MA<br>We report here that BID, a BH3 domain-containing proapoptotic Bcl2 family member, is a specific proximal substrate of Casp8 in the Fas apoptotic signaling pathway. While full-length BID is localized in cytosol, truncated BID (tBID) translocates to mitochondria and thus transduces apoptotic signals from cytoplasmic membrane to mitochondria. tBID induces first the clustering of mitochondria around the nuclei and release of cytochrome c independent of caspase activity, and then the loss of mitochondrial membrance potential, cell shrinkage, and nuclear condensation in a caspase-dependent fashion. Coexpression of BcixL inhibits all the apoptotic changes induced by tBID. Our results indicate that BID is a mediator of mitochondrial damage induced by Casp8. |
| **Bid, a Bcl2 Interacting Protein, Mediates Cytochrome c Release from Mitochondria in Response to Activation of Cell Surface Death Receptors**<br>Luo, Budihardjo, Zou, Slaughter, and Wang at the University of Texas Southwestern Medical Center, Dallas TX<br>We report here the purification of a cytosolic protein that induces cytochrome c release from mitochondria in response to caspase-8, the apical caspace activated by cell suface death receptors such as Fas and TNF. Peptide mass fingerprinting identified this protein as Bid, a BH3 domain-containing protein know to interact with both Bcl2 and Bax. Caspase-8 cleaves Bid, and the COOH-terminal part translocates to mitochondria where it triggers cytochrome c release. Immuno-depletion of Bid from cell extracts eliminated the cytochrome c releasing activity. The cytochrome c releasing activity of Bid was antagonized by Bcl2. A mutation at the BH3 domain diminished its cytochrome c releasing activity. Bid, therefore, relays an apoptotic signal from the cell surface to mitochondria. |

Both papers report the discovery of an important molecule involved in the cell death or apoptosis. The two teams found that after activation of the death receptors on the cell membrane, the death signal is carried to the mitochondria by a cytosolic protein called BID. Confirming that these two papers truly report the same scientific discovery, an August 21 2000 article in The Scientist notes that "[t]hese two Cell papers outline two independent identifications of a critical missing link in [the apoptosis] signaling pathway" (Halim 2000). As occurs frequently in the case of simultaneous discoveries, both papers were published back-to-back (pages 481-490 and 491-501) in the same issue of the same journal. As noted below, editors receiving manuscripts that report the same (or very similar) findings around the same time frequently elect to publish them back-to-back in order to underscore the reliability of important discoveries.

Of the two papers reporting the BID protein, the paper located in Boston, where local firms perform R&D in similar fields, received more references from patents than did the paper in Dallas, which is largely isolated from relevant industry. The algorithm for finding simultaneous discoveries is detailed in (Bikard 2012).

D. Measuring the Flow of Academic Science to Industry

Tracking the flow of academic science to industry empirically is challenging because such flows can take a variety of forms: licensing, consulting, strategic partnerships, and others(Roach and Cohen 2013). In a landmark paper, Jaffe, Henderson and Trajtenberg (1993) proposed that patent citations can be used to measure knowledge flow, but this measure is not without important limitations. First, patent citations have legal implications since they delimit the scope of an invention. The doctrine of "Inequitable Conduct" means that omission of information material to patentability can lead to the invalidation of the patent. Patent citations are therefore often added by patent attorney and patent examiners (Alcácer and Gittelman 2006; Alcácer, Gittelman, and Sampat 2009) and they can be used strategically (Lampe 2012). Second, each patent is by definition unique, making interpretation of non-citation difficult. Concerns regarding the definition of a control group of non-citing patents has led to major debates in the literature studying the localization of knowledge spillovers (Thompson and Fox-Kean 2005; Henderson, Jaffe, and Trajtenberg 2005). Third, knowledge is often not patented and therefore not observable using this type of measure (Griliches 1990).

Characteristics of our empirical setting allow us to largely address all three challenges. First, there is no legal requirement to refer to every paper "twin" disclosing the same simultaneous discovery. According to USPTO Rule 56 (37 CFR 1.56): "information is material to patentability when it is not cumulative to information already of record or being made of record in the application." In other words, if multiple papers disclose the same knowledge, referring to one of the papers is sufficient. References in our setting are therefore much less likely to be driven by legal or strategic considerations. Second, while every patent is by definition unique, the same is not true for scientific publications. The patent system does not recognize "ties" in the race for priority and simultaneous or independent inventions are therefore the topic of important debates in the legal literature (Vermont 2006; Lemley 2007). The same is not true in science. As we described above, when two papers make the same discovery and send it for publication at around the same time, multiple papers can be published disclosing very similar knowledge (e.g., Cozzens 1989). Third, since most academic knowledge gets published rather than patented, we focus on nonpatent references in patents.

Our focus on simultaneous discoveries is advantageous in that it allows us to measure instances in which citation was possible but did not happen. In other words, the quality of the match between the treated and the control means that we can be confident that non-referencing in our data is not driven by unobserved hetoregeneity. However, the quality of the match also comes at a cost in that inventors presumably need knowledge to flow from only one of the co-discoveries. Pessimists might argue that the fact that the discoveries are the same means that we might simply be measuring inventors' preference to reference discoveries made in specific locations. In other words, citation and non-citations in patents might not constitute a valid "trail" of knowledge flow. On the other hand, others might follow Merton (1961) and argue that the publication of simultaneous discoveries adds considerable visibility to a discovery. As a result, our approach might constitute a very conservative test of the relationship between academic location and the existence of frictions in the flow of knowledge.

Unlike patent citations to other patents, patent references to scientific publications are not readily available through existing databases. Each patent contains a list of non-patent references in the "Other References" section of a patent, which are provided as unstructured text strings. We use four easily matched criteria: 1) the surname of the first author, 2) the year of the journal, 3) volume number of journal, and 4) the starting page number. This tuple is highly unlikely to be non-unique; in order for this to occur, two authors with the same surname would have had to publish articles in different journals that had the same volume number in the same year; moreover both articles would have to start on the same page. We use the matches produced from these four criteria as a first pass to create a superset of possible matches and then inspect those by hand for Type II errors. This exercise produces our dependent variable REFERENCED.

E. Linking Simultaneous Discoveries and Patents

We apply the above methodology to our 578 simultaneous discoveries published in 1,246 papers. Given our interest in the flow of academic research to industry, we drop references from patents assigned to universities or other academic entities (although we will later use these as a placebo test to establish that location outside of an R&D hub specifically affects the flow of knowledge from academia to industry and not just knowledge flow more generally). We then eliminate self-references in two ways. First, if the surname and first initial of any author on the paper matches any inventor on the patent, we remove the paper-patent dyad from consideration. Second, and perhaps more importantly, we manually reviewed the acknowledgments section of each paper and then removed paper-patent dyads where the patent assignee was acknowledged as a sponsor of that research. Applying these restrictions reduces the sample to 313 papers reporting 146 simultaneous discoveries, with 1,910 scientific references from 1,281 patents.

As the objective of our identification strategy is to compare the likelihood of differently-located simultaneous discoveries having been referenced by patents, we then construct a dataset where every patent that referenced any of our 313 papers is paired with all papers that disclose that same simultaneous discovery. For example, given a pair of "twin" papers where one of the papers is referenced by a later patent, we also create an observation for that same patent together with the twin paper that was not referenced but could have been, given that the twin papers disclose the same simultaneous discovery. Figure 1 illustrates the setup.

Figure 1 about here

For each paper-patent dyad representing a (potential) scientific reference, we account for both temporal and spatial separation between the paper and patent. Given that our key explanatory variable reports whether a paper is not located within a hub of relevant R&D activity, the distance between paper and potentially-referencing patent is perhaps our most important control. We control for distance in a linear fashion with the logged count of miles between the paper and potentially-referencing patent: DISTANCE. Following prior research, however, we recognize that the relationship between distance and diffusion may not be linear. And thus introduce dummy variables for separation of 0-20 miles, 20-50, 50-250, 250-1000, and 1000-2500 (separation of greater than 2500 miles is the omitted category) as well as whether the paper and patent are in the SAME_COUNTRY. Lastly, twin papers are usually but not always published in the same year, so we control for the TIMELAG between the publication of the twin and the potentially-referencing patent.

At the level of the paper, we account for whether the paper is published in the U.S. (PAPER_US) given that our patents, while worldwide, are granted by the U.S. Patent & Trademark Office. We also control for the journal impact factor of the journal in which the focal paper was published (PAPER_JIF), as more prestigious journals may naturally attract more attention. Whether the focal paper reporting a simultaneous discovery was patented by its authors (PAPER_PATENTED) is an essential control as doing so places the discovery in the domain of those who submit and examine patents, forming a "patent paper pair" (Murray 2002). Whether the focal academic paper was patented may also reflect the proclivity of the paper's authors to engage in commercial activity.

To account further for heterogeneity among authors, we introduce two additional controls. AUTHOR_PAT_STOCK is the logged count of patents that had been awarded to the corresponding author of the focal paper as of the time of its publication. AUTHOR_PUB_STOCK reports a similar measure for the number of academic publications (in any journal). Both are logged due to skew.

Finally, we introduce a set of controls for the organization or institution with which the corresponding author is affiliated. The institution's publishing productivity—and resulting prestige in the academic community—is reflected by the ORG_ PRESTIGE. It is a logged count of publications that the institution has in the top 15 scientific journals between 2000 and 2010. A series of controls also account for the institution's general proclivity to engage in translational activities. ORG_STOCK_PATS is the log of patents granted to the institution in the past five years. ORG_ INDUSTRY_FUNDING is the logged R&D dollars provided by firms to the institution and may represent the closeness of that institution to relevant (but not necessarily local) industry. Table 1 summarizes the variables in the paper. Descriptive statistics are in Table 2.

Tables 1 and 2 about here

F. Measuring the Location of Relevant Hubs of Industrial R&D

To measure weather an academic institution is located inside or outside a relevant hubs of industrial R&D, we focus on inventive activity (a) in the relevant field (b) within 5 years of the discovery and (c) within a specific radius of the institution. This measure is operationalized as follows. We start by collecting the technological subclassifications from all patents that contain scientific references to one of the 313 "twin" papers in our sample. Patents referencing papers that report the simultaneous discoveries are categorized into 712 unique subclasses. For each subclass, we then collect all non-university patents belonging to that subclass, whether or not they reference any of the twins in our study. We find a total of 1,430,822 corporate patents that were categorized by the USPTO into one of the 712 technology subclasses.

We then construct "hubs" of industrial R&D activity as follows. For each of the 712 technology subclasses that characterize our simultaneous discoveries, we collect the locations in which those non-university patents are found in that subclass. For each location, we count the number of patents in that same subclass within a 50-mile radius for each half-decade. We divide those two figures to yield the percentage of overall patenting activity from that technology subclass occurring in that location. We label a location as a "hub" of industrial R&D for that subclass if more than 5% of patents in that technology subclass are located within a 50-mile radius. Because this threshold can easily be exceeded in technology subclasses with very few patents (e.g., in a subclass with only 20 patents, every location has at least 5% of patenting), we require that a location have at least five patents in that subclass to qualify as a "hub." This exercise yields a list of R&D hubs for each of the 712 technology subclasses relevant to our simultaneous discoveries within five years of the publication date. (Some subclasses are widely distributed across locations and thus do not have any hubs.)

To determine whether a given academic paper is inside or outside of a relevant hub of industrial R&D, we first make a list of the technological subclasses for all patents that referenced either the focal paper or any of its twins. These patent subclasses delimit the relevant scope of R&D activity for that simultaneous discovery. For each twin paper reporting that simultaneous discovery, we then check whether there is at least one R&D hub within 50 miles (i.e., commuting distance) of the corresponding address of the focal paper (likely the location of the lab where the research was conducted). If we cannot find a hub within 50 miles, we set OUTSIDE_R&D_HUBS to 1 for that paper.

To illustrate the concept of being located inside or outside of a hub of relevant industrial R&D, we return to our simultaneous discovery from above. Again, we examine two papers in the August 1998 issue of Cell, one at Harvard Medical School in Boston, MA and another at UT Southwestern Medical Center in Dallas, TX. In determining whether either of these research teams was inside or outside of the relevant R&D hubs, we first note that 19 patents (either firm-owned or university-owned[1]) listed one of these papers as a scientific reference.[2] We then define the scope of relevant R&D by obtaining the USPTO technological subclassifications for these patents. A few have the same classification, yielding 17 subclasses.

The next step is to locate "hubs" of industry R&D in these technological areas. We find 3858 firm-owned patents that were assigned to these subclasses during 1995-1999 (again, the article was published in 1998). The locations with R&D "hubs" containing at least five patents and more than 5% of patenting activity for the above 17 subclasses include Milan, Italy; La Jolla, Santa Clara, and Solana Beach, California; Canton, Lexington, and Weston, Massachusetts; Chevy Chase and Silver Spring, Maryland; Berkeley Heights, Old Bridge, and Teaneck, New Jersey, and Bainbridge Island, Washington.

To determine whether Harvard Medical School and UT Southwestern Medical Center are located inside or outside of a relevant R&D hub, we then check whether either institution is within commuting distance (i.e., 50 miles) of those cities. While Harvard Medical School in Boston is within commuting distance of Canton, Lexington, and Weston Massachusetts, Dallas is far from any of the cities listed above as having at least five patents and more than 5% of patents in the subclass. Thus we classify the

[1] We use both industry and academic patents to compile the list of relevant technology subclasses, but our measure of knowledge flowing from academia to industry includes only references from firm-assigned patents.

[2] Both papers were referenced by three patents (7452869, 7638324, and 7745109). The Dallas paper was referenced exclusively by three patents: 6503754, 7247700, and 7829662. The Boston paper The Boston paper was referenced exclusively by 13 patents: 6221355, 6245885, 6326354, 6645501, 6692927, 6773911, 6946458, 7026472, 7371834, 7381713, 7514413, 7635693, and 7772202.

paper published in Dallas as lying outside a hub of relevant industrial R&D and that published in Boston as lying inside such a hub.

Applying this definition, 67.6% of our paper twins are outside of a hub of relevant industrial R&D. Our definition of R&D hubs is somewhat conservative, requiring only 5% of patents in the relevant subclass. Less conservative formulations (e.g., requiring more than 10% of patenting in the subclass), yield similar results and label approximately 80% of papers as lying outside the relevant R&D hubs.

G. Empirical Setup

We examine the impact of the geographic location of academic institutions on the flow of public research to industrial R&D by examining the references of academic papers disclosing the same discovery in corporate patents. An observation is a dyad of a published paper reporting a simultaneous discovery and a patent that is at risk of referencing the paper as non-patent reference. Our analysis leverages the simultaneous-discovery nature of our data since a patent that references one paper is presumably at a similar risk of referencing any of its "twins" as described in Figure 1. We specify a conditional logit model with fixed effects for the combination of a group of paper twins reporting a simultaneous discovery and a patent that references one but not all of the twins.[3] Thus our analysis reveals whether, for a given patent that in theory could reference any of the twin papers reporting the simultaneous discovery, the factors associated with one of that set of twin papers being referenced. The regression equation is given as

$$REFERENCED_{ijk} = f(\varepsilon_{ijk};\ \alpha_0 + \alpha_1 OUTSIDE\_R\&D\_HUBS_i + \alpha_{2^\circ}\bar{X}_{ik} + \gamma_{jk})$$

where j represents the simultaneous discovery, i represents the paper reporting the simultaneous discovery, and k represents the potentially-referencing patent. $OUTSIDE\_R\&D\_HUBS_i$ is our main explanatory variable and is defined at the paper-patent dyad level as described in section 2.E. $\gamma_{jk}$ is our simultaneous-discovery/patent fixed effect. Finally, $X_{ik}$ is a vector of covariates including the geographic distance between the focal paper and the potentially-referencing patent. Some specifications also include city and academic institution-level fixed effects. Standard errors are clustered at the level of the simultaneous discovery.

## 3. Empirical Results

### A. Replication and basic results

We begin our analysis in Table 3. Column (1) explores the influence of control variables. As one might expect, the paper has been patented (PAPER_PATENTED) somewhat affects the likelihood of being

---

[3] Results are robust to a linear-probability specification and also to specifying fixed effects just at the level of the simultaneous discovery as opposed to the simultaneous-discovery/patent level.

referenced, possibly an indicator of commercially oriented efforts on behalf of the scientists or their institution. Journal impact factor (PAPER_JIF) is not impactful among our articles, perhaps because we sample on simultaneous discoveries that are highly cited in the academic literature. Papers with authors who have more patents and papers may be more likely to be referenced, but the coefficient is noisy. We do not see strong impact of organizational characteristics on the likelihood of being referenced. Unsurprisingly, twin papers published earlier (and thus with a longer TIMELAG until the focal patent was submitted) are somewhat more likely to receive a reference, although the coefficient is noisy.

Table 3 about here

In column (2) we introduce geographic location to our analysis. Consistent with prior work on the geographic localization of knowledge diffusion, DISTANCE is negatively associated with the likelihood of a paper being referenced by a patent. That our analysis of simultaneous discoveries yields similar localization dynamics as seen in prior diffusion studies helps to allay concerns that this set of 313 papers might exhibit strongly different characteristics than larger samples analyzed previously (but without the benefit of accounting for the quality of the science). Following Singh and Marx (2013), in column (3) we switch from parametrically modeling distance to a non-parametric series of dummies in order to capture the nuances of localization. Compared to an omitted category of more than 2,500 miles separating the patent and paper, a reference is substantially more likely to occur only when the patent and paper are within 20 miles of each other. Moreover, the magnitude of the effect is stark: the average marginal effect on being referenced when a paper twin when the two are within ten miles of each other is 36.7%. This sharp dropoff with even a small separation between patent and paper resembles the Belenzon and Schankerman (2013) finding that the probability of a paper being cited by a patent drops by 40% when they are separated by more than 25 miles. The minor differences may be attributable to our sampling methodology, which analyzes a smaller number of academic papers but enables selection of “twin” controls. Again, our ability to replicate prior results regarding localization lends support to the notion that valid inferences can be drawn from studying this admittedly small sample of discoveries.

In column (4) of Table 3, we move from analyzing the distance between the source of the academic knowledge and the potential firm (as is common in prior literature) to examining the implications of an academic scientist operating outside of “hubs” of relevant industrial R&D. In doing so we continue to control for spatial separation between paper and patent, as a key concern might be that if far-flung academics are not inside hubs of R&D they might also be far from firms that would want to use their knowledge. Adding OUTSIDE_R&D_HUBS does not substantially influence the coefficients on the distance dummies from column (3), but its own coefficient is negative with statistical significance at the 1% level. Paper twins located outside of R&D hubs are 9.97% less likely to be cited than their twin(s) inside R&D hubs, but this average marginal effect is obtained almost entirely among patent-paper dyads that are not located within ten miles of each other. Thus it appears that the deleterious effect of being separated from a potential user of academic knowledge can be ameliorated if the knowledge producer is co-located with communities of commercial inventors in the same field. To return to our example from earlier, although we might expect a firm to be less likely to build upon an academic discovery 1,000 miles away when an equivalent discovery is only 100 miles away, the more distant discovery may in fact be more likely to be referenced if it is within a hub of industrial R&D in that field.

### B. Robustness and placebo tests

Table 4 presents our robustness analysis. Column (1) introduces quadratics for the impact factor of the paper's journal, the corresponding author's stock of papers and patents, as well as the organization's prestige and stock of patents. Results are similar as before.

Column (2) shows just one of a series of leave-one-out tests to address the concern that our results might depend on a high concentration of academic papers in a particular city. Boston Massachusetts is home to 12.5% of our twin papers, yet the coefficient on OUTSIDE_R&D_HUBS is robust to its omission. We achieve similar results when excluding sequentially each city with more than 5% of our twin papers: San Diego, New York, Bethesda Maryland, and Toronto. In unreported results, we repeat the leave-one-out test for academic institutions with a larger percentage of papers (Harvard University being the largest, with 6.7%) and also for patent assignees (Genentech the largest, with 5.25%). Thus our results do not depend on any one city, academic institution, or patent assignee.

An additional test of sensitivity to locational characteristics is conducted in column (3). One might wonder whether hubs of R&D activity are co-located and thus correlated with other locational characteristics that might also explain the apparent effect of hubs on the flow of academic knowledge to industry. Assume for the sake of argument that hubs of R&D tend to be in the same cities as prominent universities. Then it might not be that the academic article was referenced because it was published within a hub of industrial R&D but rather because it is located in a city with reputable academic institutions that firms tend to pay attention to in the same way that they might be more aware of a more prestigious institution. We create a variable CITY_ PRESTIGE which captures the number of publications in top-15 scientific journals that occurred at any institution located that city between 2000 and 2010. (CITY_ PRESTIGE and ORG_PRESTIGE are rather highly correlated, so the latter is excluded from column (3).) As is visible in column (3), no statistically-significant association is visible, so it does not appear that concentration of academic activity in a city serves as an alternative explanation. In unreported results, we also find no effect when controlling for the population of each institution's city.

In column (4) we test our assumption that a hub of industrial activity is defined as a location with at least 5% of patenting activity in that technological subclass. It is quite unusual to find locations with a fifth or a quarter of patenting, so we do not test 20% or 25% thresholds, but in column (4) we double the percentage required to 10%. Results are consistent with those obtained at the 5% level (indeed, all results in all tables can be recovered with this different definition).

Column (5) replaces the explanatory variable with one in which location outside an R&D hub is expanded to include any author; if even one of the authors is located in a hub of relevant industry R&D, the paper is not considered to lie outside of an R&D hub. Doing so reduces the percentage of papers classified as outside of relevant R&D hubs from 67.6% to 59.6%, as shown in Table 2. Results are quite similar (indeed, all results in all tables can be recovered with this adjusted explanatory variable).

Finally, in column (6) we perform a placebo test. The analyses so far measure references to non-firm-authored papers from firm-assigned patents in order to measure the flow of knowledge from academia to industry. If location outside an R&D hub affects only these flows and not knowledge diffusion more generally, the geography of industrial R&D should not affect flows within academia, i.e., references to academic papers from university-owned patents. (Note: we are not measuring here universities patenting the discovery in question—captured by the PAPER_PATENTED—variable, but rather references to the academic paper that appear in university-owned patents not including those filed by the authors of the paper.) In column (6) we construct dyads of paper twins and university-assigned patents that reference them, excluding as above self-references by either the authors of the paper or other scientists at the same institution. While the sign of the coefficient on OUTSIDE_R&D_HUBS is negative, it fails to achieve statistical significance at conventional levels. As expected, considering our method for identifying simultaneous discoveries, location inside or outside of a hub of industrial R&D does not appear to materially impact the flow of academic discoveries to university patents.

Table 4 about here

### C. Mechanisms

In Table 5 we further explore the mechanisms underlying the attenuation of knowledge flow to industry from academics located outside of relevant R&D hubs. Our first potential moderator concerns the relationship between the institution and industry. If proximity to relevant R&D hubs facilitates the flow of knowledge from academia to industry as informal interactions are more likely to arise in close proximity, then more formal relationships between academia and industry may act as a substitute. We explore this interaction in column (1) of Table 5 by introducing a variable that measures industry investment in R&D at the institution where a given paper was published (ORG_INDUSTRY_FUNDING). Given that this data is collected from the Association of University Technology Managers, our analysis in column (1) is necessarily limited to North American institutions. The positive coefficient on the interaction of ORG_INDUSTRY_FUNDING and the indicator for being outside of relevant R&D hubs is indeed suggestive of a substitution effect, but in column (2) we decompose the mechanism further.

Table 5 about here

In column (2) we replace the interaction term between R&D hubs and industry investment in R&D at a focal institution with a set of indicators interacting whether the paper is outside an R&D hub with four levels of industry funding of R&D at the institution: NO_FUNDING, LOWER_FUNDING, HIGHER_FUNDING, and HIGHEST_FUNDING. The coefficients suggest that the apparent substitution effect in column (1) is driven primarily by the fact that papers written at institutions that are both not near relevant R&D hubs and which also do not have any industry funding are much less likely to be referenced. Given the challenges of interpreting interaction terms in nonlinear models such as conditional logit, we repeat this analysis with a linear probability model and graph the resulting coefficients in Figure 2. Panel A of Figure 2 corresponds to column (2) of Table 5. Coefficients appear to be monotonically decreasing as the level of funding grows, but the only coefficient significantly different from zero is for papers that are outside of relevant R&D hubs and also receiving no R&D investment dollars from industry. Thus it appears that the negative effect of being located outside of R&D hubs is felt most acutely along the margin of institutions that lack formal connections to industry more generally.

Figure 2 about here

In column (3), we consider the possibility that institutional reputation or prestige might substitute for proximity to relevant hubs. Simply put, papers written at institutions that are more prominent may be referenced disproportionately often and thus less affected by being located outside hubs of relevant R&D. As in column (2), we interact the OUTSIDE_R&D_HUBS indicator with indicators for the four quartiles of institutional prestige. (The omitted category is HIGHEST_PRESTIGE.) Again relying on the graphed linear-probability coefficients for inference, Panel B of Figure 2 indicates that papers located outside of relevant R&D hubs and whose institutions are in the lowest quartile of prestige are least likely to be referenced. The other coefficients are smaller and similar in magnitude, with one category weakly distinct from zero. Thus it appears that the negative effect of being located outside of R&D hubs is felt most acutely along the margin of low-prestige institutions.

Finally, we examine the interplay between separation from R&D hubs and the distance between the paper and the potentially-referencing patent. As in the previous two columns, we add interaction terms with all distance dummies, including for the omitted category of more than 2500 miles separating the two. (As before, the base variables are not shown to conserve space in Table 5.) The pattern revealed by the plot of linear-probability coefficients in Panel C of Figure 2 suggests a three-tiered effect of separation from relevant R&D hubs. For paper-patent dyads within commuting distance of each other (i.e., either less than 20 miles or 20-50 miles), being isolated from relevant R&D hubs appears not to further discourage referencing. Given that paper-patent dyads within 20 miles of each other are much more likely to be referenced, it appears that being outside an R&D hub may not matter if the potentially-citing patent is close enough to the paper. R&D hubs also do not appear to influence the probability of a paper-patent dyad containing a reference when the two are separated by extreme distances. The negative impact of location outside a hub of relevant R&D appears particularly salient however for paper-patent dyads where separation is more than 50 but fewer than 2500 miles. In other words, we find evidence that location outside of a relevant R&D hub appears hinders knowledge flow unless those inventors work very close or very far from the academic institution in which the discovery was made.

## 4. Discussion

Our analysis of simultaneous discoveries shows that the publication of new scientific knowledge alone does not guarantee that industrial inventors will use it. Inventors in firms are unlikely to have perfect visibility into the latest academic developments and might therefore overlook valuable discoveries. The geographic location of academic research institutions can be a source of such frictions in knowledge flow because the geography of academic research does not perfectly coincide with that of industrial R&D. Here, we measure the flow (or non-flow) of science by observing patent referencing (or not) academic publications disclosing the same discovery but that emerged in different locations. We are able to replicate prior findings regarding the localization of knowledge spillovers. However, our analysis focuses on another aspect of geographic location that has received much less attention: whether an academic institution is based in a hub of relevant industrial R&D or outside of it.

Our findings bear directly on public policy regarding the dissemination of academic science. The current distribution of academic research organization might promote equal access to science across geographical areas, but our results suggest that such efforts toward egalitarianism may come at a cost, by systematically complicating firms' exploitation of discoveries made at remote academic institutions. The variance that we observe in the flow of academic science to industry presents therefore something of a dilemma for science policy. If funding to institutions located outside the relevant R&D hubs is in fact less efficient in terms of producing discoveries that are impactful outside of academia, is it rational—and perhaps welfare-enhancing—to purposely channel funding away from those institutions? Or should policy-makers introduce measures to specifically promote the dissemination of scientific knowledge produced at those institutions so that their results are not neglected by firms?

## References


Adams, James D. 1990. “Fundamental Stocks of Knowledge and Productivity Growth.” Journal of Political Economy 98 (4): 673–702.

———. 2002. “Comparative Localization of Academic and Industrial Spillovers.” Journal of Economic Geography 2 (3): 253–78.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. “Academic Freedom, Private-Sector Focus, and the Process of Innovation.” The RAND Journal of Economics 39 (3): 617–35.

Agrawal, A., and A. Goldfarb. 2008. “Restructuring Research: Communication Costs and the Democratization of University Innovation.” The American Economic Review 98 (4): 1578.

Agrawal, Ajay, and Rebecca M. Henderson. 2002. “Putting Patents in Context: Exploring Knowledge Transfer from MIT.” Management Science 48 (1).

Alcácer, Juan, and Michelle Gittelman. 2006. “Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations.” Review of Economics and Statistics 88 (4): 774–79. doi:i: 10.1162/rest.88.4.774</p>.

Alcácer, Juan, Michelle Gittelman, and Bhaven Sampat. 2009. “Applicant and Examiner Citations in U.S. Patents: An Overview and Analysis.” Research Policy 38 (2): 415–27. doi:16/j.respol.2008.12.001.

Audretsch, David B., and Paula E. Stephan. 1996. “Company-Scientist Locational Links: The Case of Biotechnology.” The American Economic Review 86 (3): 641–52.

Azoulay, Pierre, Joshua S. Graff Zivin, and Bhaven N. Sampat. 2012. “The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine.” In The Rate of Direction of Inventive Activity, 107–55. University of Chicago Press. http://www.nber.org/papers/w16683.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. 2010. “Superstar Extinction.” The Quarterly Journal of Economics 125 (2): 549–89.

Babbage, Charles. 1832. On the Economy of Machinery and Manufactures ... Second Edition Enlarged. Charles Knight.

Belenzon, Sharon, and Mark Schankerman. 2013. “Spreading the Word: Geography, Policy, and Knowledge Spillovers.” Review of Economics and Statistics 95 (3): 884–903. doi:10.1162/REST_a_00334.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2002. “Links and Impacts: The Influence of Public Research on Industrial R&D.” Management Science 48 (1): 1–23.

Cozzens, Susan E. 1989. Social Control and Multiple Discovery in Science: The Opiate Receptor Case. State University of New York Press.

Dasgupta, Partha, and Paul A. David. 1994. "Toward a New Economics of Science." Research Policy 23 (5): 487–521.

David, Paul A. 2001. "From Keeping 'Nature's Secrets' to the Institutionalization of 'Open Science.'" Working Paper 01006. Stanford University, Department of Economics. https://ideas.repec.org/p/wop/stanec/01006.html.

Drahl, Carmel. 2014. "Consecutive Journal Publications Illuminate Collaboration And Compromise In Chemistry." Chemical & Engineering News 92 (40). http://cen.acs.org/articles/92/i40/Consecutive-Journal-Publications-Illuminate-Collaboration.html.

Furman, Jeffrey, and Megan J. MacGarvie. 2007. "Academic Science and the Birth of Industrial Research Laboratories in the U.S. Pharmaceutical Industry." Journal of Economic Behavior & Organization 63 (4): 756–76. doi:10.1016/j.jebo.2006.05.014.

Furman, Jeffrey, and Scott Stern. 2011. "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research." American Economic Review 101 (5): 1933–63.

Galasso, Alberto, and Mark Schankerman. 2014. "Patents and Cumulative Innovation: Causal Evidence from the Courts*." The Quarterly Journal of Economics, November, qju029. doi:10.1093/qje/qju029.

Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." Journal of Economic Literature 28 (4): 1661–1707. doi:10.2307/2727442.

———. 1998. "Introduction to 'R&D and Productivity: The Econometric Evidence.'" In R&D and Productivity: The Econometric Evidence, 1–14. University of Chicago Press. http://www.nber.org/books/gril98-1.

Grossman, Gene M., and Elhanan Helpman. 1993. Innovation and Growth in the Global Economy. MIT Press.

Halim, Nadia. 2000. "Bridging Apoptotic Signaling Gaps." The Scientist, August 21. http://www.the-scientist.com/?articles.view/articleNo/12974/title/Bridging-Apoptotic-Signaling-Gaps/.

Harris, Gardiner. 2011. "Federal Research Center Will Help Develop Medicines." The New York Times, January 22, sec. Health / Money & Policy. http://www.nytimes.com/2011/01/23/health/policy/23drug.html.

Henderson, Rebecca M., Adam B. Jaffe, and Manuel Trajtenberg. 1998. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988." Review of Economics and Statistics 80 (1): 119–27. doi:10.1162/003465398557221.

———. 2005. “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment.” The American Economic Review 95 (1): 461–64.

Jaffe, Adam B. 1989. “Real Effects of Academic Research.” The American Economic Review 79 (5): 957–70.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca M. Henderson. 1993. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” The Quarterly Journal of Economics 108 (3): 577–98. doi:10.2307/2118401.

Lampe, Ryan. 2012. “Strategic Citation.” Review of Economics and Statistics 94 (1): 320–33. doi:10.1162/REST_a_00159.

Lemley, Mark A. 2007. “Should Patent Infringement Require Proof of Copying?” Michigan Law Review 105 (7): 1525–36.

Li, Guan-Cheng, Ronald Lai, Alexander D’Amour, David M. Doolin, Ye Sun, Vetle I. Torvik, Amy Z. Yu, and Lee Fleming. 2014. “Disambiguation and Co-Authorship Networks of the U.S. Patent Inventor Database (1975–2010).” Research Policy 43 (6): 941–55. doi:10.1016/j.respol.2014.01.012.

Mansfield, Edwin. 1998. “Academic Research and Industrial Innovation: An Update of Empirical Findings.” Research Policy 26 (7-8): 773–76. doi:10.1016/S0048-7333(97)00043-7.

Marshall, A. 1895. Principles of Economics. Macmillan.

Merton, Robert K. 1973. The Sociology of Science: Theoretical and Empirical Investigations. University of Chicago Press.

Mokyr, Joel. 2002. The Gifts of Athena: Historical Origins of the Knowledge Economy. Princeton University Press.

Murray, Fiona. 2002. “Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering.” Research Policy 31 (8-9): 1389–1403.

Murray, Fiona, and Scott Stern. 2007. “Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?: An Empirical Test of the Anti-Commons Hypothesis.” Journal of Economic Behavior & Organization 63 (4): 648–87. doi:10.1016/j.jebo.2006.05.017.

Nelson, Richard R. 1959. “The Simple Economics of Basic Scientific Research.” Journal of Political Economy 67 (3): 297–306.

Nelson, Richard R. 1982. “The Role of Knowledge in R&D Efficiency.” The Quarterly Journal of Economics, The Quarterly Journal of Economics, 97 (3): 453–70.

Niehans, Jurg. 1995. “Multiple Discoveries in Economic Theory.” European Journal of the History of Economic Thought 2 (1): 1. doi:Article.

Ogburn, William F., and Dorothy Thomas. 1922. "Are Inventions Inevitable? A Note on Social Evolution." Political Science Quarterly 37 (1): 83–98.

Roach, Michael, and Wesley M. Cohen. 2013. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." Management Science 59 (2): 504–25. doi:10.1287/mnsc.1120.1644.

Romer, Paul M. 1990. "Endogenous Technological Change." Journal of Political Economy 98 (5 pt 2). http://www.dklevine.com/archive/refs42135.pdf.

Rosenberg, Nathan. 1994. Exploring the Black Box: Technology, Economics, and History. Cambridge University Press.

Rosenberg, Nathan, and Richard R. Nelson. 1994. "American Universities and Technical Advance in Industry." Research Policy 23 (3): 323–48. doi:10.1016/0048-7333(94)90042-6.

Singh, Jasjit, and Matt Marx. 2013. "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity." Management Science 59 (9): 2056–78. doi:10.1287/mnsc.1120.1700.

Sohn, Eunhee. 2014. "The Endogeneity of Academic Science to Local Industrial R&D." Academy of Management Proceedings 2014 (1): 11413. doi:10.5465/AMBPP.2014.286.

Stephan, Paula E. 1996. "The Economics of Science." Journal of Economic Literature 34 (3): 1199–1235.

Stigler, George J. 1980. "Merton on Multiples, Denied and Affirmed." Transactions of the New York Academy of Sciences 39 (1 Series II): 143–46. doi:10.1111/j.2164-0947.1980.tb02774.x.

Thompson, Peter, and Melanie Fox-Kean. 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." The American Economic Review 95 (1): 450–60.

Vermont, Samson. 2006. "Independent Invention as a Defense to Patent Infringement." Michigan Law Review 105 (3): 475–504.

Weintraub, E. Roy. 2011. "Retrospectives: Lionel W. McKenzie and the Proof of the Existence of a Competitive Equilibrium." Journal of Economic Perspectives 25 (2): 199–215. doi:10.1257/jep.25.2.199.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." The American Economic Review 88 (1): 290–306.

**Table 1:** Variable definitions.

| | |
|---|---|
| paper-level | |
| OUTSIDE_R&D_HUBS | The academic institution in which the paper's corresponding author is based is not within 50 miles (commuting distance) of any city with at least five patents and more than 2% of patenting in relevant technological subclasses. |
| ALL_OUTSIDE_R&D_HUBS | Same as OUTSIDE_R&D_HUBS except that none of the authors is within 50 miles of a hub of relevant industrial R&D |
| PAPER_JIF | Impact factor of journal paper was published in, calculated as five-year running average as of 2009 (logged). |
| PAPER_US | Paper's corresponding address is in U.S. |
| PAPER_PATENTED | One of paper's authors was granted a patent on the discovery reported by the paper. |
| author-level | |
| AUTHOR_PAT_STOCK | Number of patents the corresponding author has previously filed. |
| AUTHOR_PUB_STOCK | Number of papers the corresponding author had previously published. |
| organization-level | |
| ORG_PAT_STOCK | Log number of patents filed by this institution in the past 5 years. |
| ORG_PRESTIGE | Log number of papers published by this institution in the top 15 scientific journals between 2000 and 2010. |
| ORG_LOWEST_PRESTIGE | Institution is in the bottom (4th) quartile of publications in the top 15 scientific journals between 2000 and 2010 (i.e., fewer than 10). |
| ORG_LOWER_PRESTIGE | Institution is in the second-lowest (3rd) quartile of publications in the top 15 scientific journals between 2000 and 2010 (i.e., 11-40). |
| ORG_HIGHER_PRESTIGE | Institution is in the second-highest (2nd) quartile of publications in the top 15 scientific journals between 2000 and 2010 (i.e., 41-125). |
| ORG_HIGHEST_PRESTIGE | Institution is in the top (1st) quartile of publications in the top 15 scientific journals between 2000 and 2010 (i.e., more than 125). |
| ORG_INDUSTRY_FUNDING | Log dollars invested in the institution's research by outside firms. |
| NO_INDUSTRY_FUNDING | Institution received zero dollars of research funding from industry. |
| LOW_INDUSTRY_FUNDING | Institution received more than zero but less than $10MM of industry funding. |
| HIGHER_INDUSTRY_FUNDING | Institution received $10-20MM in industry funding. |
| HIGHEST_INDUSTRY_FUNDING | Industry received more than $20MM in industry funding. |
| paper/patent dyad | |
| REFERENCED | Focal paper is referenced by patent in the paper-patent dyad. |
| TIMELAG | Time lag between publication of focal paper and possibly-referencing patent. |
| DISTANCE | Log spatial distance (in miles) between paper and patent, adjusted for curvature of the earth. |
| DIST0_20 | Paper and potentially-referencing patent are within 20 miles of each other. |
| DIST20_50 | Paper and potentially-referencing patent are 20-50 miles apart. |
| DIST50_250 | Paper and potentially-referencing patent are 50-250 miles apart. |
| DIST250_1000 | Paper and potentially-referencing patent are 250-1000 miles apart. |
| DIST1000_2500 | Paper and potentially-referencing patent are 1000-2500 miles apart. |
| DIST>2500 | Paper and potentially-referencing patent are more than 2500 miles apart. |
| SAME_COUNTRY | Paper and potentially-referencing patent are in the same country |

**Table 2:** Descriptive statistics and correlations for simultaneous discoveries**.**

| | Mean | S.D. | Min | Max | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) REFERENCED | 0.477 | 0.5 | 0 | 1 | 1 | | | | | | | | | | | | |
| (2) OUTSIDE_R&D_HUBS | 0.676 | 0.47 | 0 | 1 | -0.22 | 1 | | | | | | | | | | | |
| (3) ALL_OUTSIDE_R&D_HUBS | 0.596 | 0.49 | 0 | 1 | -0.2 | 0.84 | 1 | | | | | | | | | | |
| (4) PAPER_JIF | 3.202 | 0.58 | 0 | 3.96 | -0.02 | -0.14 | -0.07 | 1 | | | | | | | | | |
| (5) PAPER_US | 0.664 | 0.47 | 0 | 1 | 0.15 | -0.46 | -0.35 | -0.04 | 1 | | | | | | | | |
| (6) PAPER_PATENTED | 0.24 | 0.43 | 0 | 1 | 0.14 | -0.1 | -0.12 | -0.06 | 0.07 | 1 | | | | | | | |
| (7) AUTHOR_PAT_STOCK | 0.439 | 0.74 | 0 | 4.33 | 0.15 | -0.18 | -0.16 | -0.15 | 0.13 | 0.16 | 1 | | | | | | |
| (8) AUTHOR_PUB_STOCK | 3.577 | 1.33 | 0 | 6.46 | 0.08 | 0 | 0.01 | -0.09 | -0.11 | 0.1 | 0.5 | 1 | | | | | |
| (9) ORG_PAT_STOCK | 3.181 | 2.31 | 0 | 7.26 | 0.09 | -0.16 | -0.11 | -0.07 | 0.6 | 0.06 | 0.11 | -0.01 | 1 | | | | |
| (10) ORG_PRESTIGE | 3.36 | 2.01 | 0 | 6.36 | 0.07 | -0.17 | -0.12 | 0.12 | 0.29 | 0.01 | -0.08 | -0.07 | 0.57 | 1 | | | |
| (11) ORG_INDUSTRY_FUNDING | 11.89 | 18.3 | 0 | 165 | 0.05 | -0.05 | -0.02 | -0.18 | 0.34 | 0.2 | 0.03 | 0.05 | 0.56 | 0.33 | 1 | | |
| (12) TIMELAG | 5.142 | 3.24 | 0 | 17 | 0.02 | -0.1 | -0.05 | 0.16 | -0.01 | -0.03 | -0.09 | -0.19 | -0.13 | 0.04 | -0.07 | 1 | |
| (13) DISTANCE | 7.167 | 1.95 | 0 | 9.26 | -0.19 | 0.27 | 0.22 | -0.09 | -0.29 | -0.12 | -0.05 | -0.01 | -0.1 | -0.18 | -0.06 | 0.04 | 1 |
| (14) SAME_COUNTRY | 0.51 | 0.5 | 0 | 1 | 0.11 | -0.35 | -0.27 | -0.03 | 0.68 | 0.02 | 0.09 | -0.04 | 0.46 | 0.25 | 0.25 | -0.07 | -0.51 |

Notes: Variables are defined in Table 1. Observations are constructed for all combinations of all twin academic papers and potentially-referencing corporate patents for all simultaneous discoveries where at least one of the twin papers is referenced by some patent. Dummies corresponding to continuous variables for DISTANCE, ORG_PRESTIGE, and ORG_INDUSTRY_FUNDING are not shown to conserve space.

**Table 3:** The impact of the location of academic institutions on industry patent referencing to papers

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| OUTSIDE_R&D_HUBS | | | | -0.832** |
| | | | | (0.296) |
| PAPER_JIF | -0.506 | -0.473 | -0.398 | -0.332 |
| | (0.380) | (0.396) | (0.359) | (0.338) |
| PAPER_US | 0.723* | 0.516 | 0.516 | 0.242 |
| | (0.333) | (0.320) | (0.440) | (0.436) |
| PAPER_PATENTED | 0.536+ | 0.368 | 0.353 | 0.328 |
| | (0.289) | (0.263) | (0.262) | (0.258) |
| AUTHOR_PAT_STOCK | 0.366 | 0.329 | 0.311 | 0.244 |
| | (0.245) | (0.221) | (0.208) | (0.189) |
| AUTHOR_PUB_STOCK | 0.0984 | 0.0790 | 0.0717 | 0.0208 |
| | (0.115) | (0.120) | (0.113) | (0.119) |
| ORG_PAT_STOCK | -0.0330 | -0.000879 | 0.00439 | 0.0251 |
| | (0.0918) | (0.0846) | (0.0814) | (0.0804) |
| ORG_PRESTIGE | 0.0664 | 0.0581 | 0.0695 | 0.0351 |
| | (0.0691) | (0.0665) | (0.0657) | (0.0664) |
| TIMELAG | 0.590+ | 0.666* | 0.671* | 0.504+ |
| | (0.325) | (0.287) | (0.284) | (0.264) |
| DISTANCE | | -0.182* | | |
| | | (0.0709) | | |
| DIST0-20 | | | 1.460* | 1.458* |
| | | | (0.651) | (0.667) |
| DIST20-50 | | | 0.143 | 0.195 |
| | | | (0.657) | (0.710) |
| DIST50-250 | | | 0.536 | 0.814 |
| | | | (0.673) | (0.630) |
| DIST250-1000 | | | 0.715* | 0.636+ |
| | | | (0.359) | (0.350) |
| DIST1000-2500 | | | 0.330 | 0.360 |
| | | | (0.345) | (0.339) |
| SAME_COUNTRY | | | -0.100 | -0.187 |
| | | | (0.580) | (0.546) |
| Observations | 1,638 | 1,638 | 1,638 | 1,638 |
| # articles | 313 | 313 | 313 | 313 |
| # twin-pat dyads | 771 | 771 | 771 | 771 |
| # simultaneous discoveries | 146 | 146 | 146 | 146 |
| Pseudo-R2 | 0.0968 | 0.122 | 0.129 | 0.149 |
| Log-likelihood | -518 | -503.3 | -499.7 | -487.8 |

Standard errors in parentheses; ** p<0.01, * p<0.05, + p<0.1

Notes: Dependent variable indicates whether a paper-patent dyad is linked by an actual reference. OUTSIDE_R&D_HUBS refers to a paper not within 50 miles of a “hub” of R&D activity for the patent subclasses associated with its simultaneous discovery. Standard errors are clustered at the level of the simultaneous discovery.

**Table 4:** The impact of the academic institution's location outside of an R&D hub: robustness

| | (1) squared terms | (2) leave-one-out (Boston) | (3) academic concentration | (4) 10% cutoff for hubs | (5) distributed authors | (6) university placebo |
|---|---|---|---|---|---|---|
| OUTSIDE_R&D_HUBS | -0.841** | -0.749* | -0.769* | -1.410** | | -0.373 |
| | (0.297) | (0.349) | (0.340) | (0.501) | | (0.299) |
| ALL_OUTSIDE_R&D_HUBS | | | | | -0.668** | |
| | | | | | (0.252) | |
| PAPER_JIF | 1.173 | -0.137 | -0.149 | -0.213 | -0.344 | 0.761** |
| | (1.357) | (0.357) | (0.365) | (0.323) | (0.342) | (0.289) |
| PAPER_US | 0.190 | 0.192 | 0.154 | 0.276 | 0.333 | -0.802+ |
| | (0.446) | (0.435) | (0.435) | (0.433) | (0.426) | (0.462) |
| PAPER_PATENTED | 0.356 | 0.332 | 0.321 | 0.215 | 0.301 | -0.243 |
| | (0.303) | (0.317) | (0.320) | (0.272) | (0.263) | (0.224) |
| AUTHOR_PAT_STOCK | 0.546 | 0.227 | 0.232 | 0.201 | 0.234 | 0.0345 |
| | (0.430) | (0.202) | (0.194) | (0.186) | (0.190) | (0.170) |
| AUTHOR_PUB_STOCK | 0.0523 | -0.0799 | -0.0860 | 0.0179 | 0.0577 | 0.0733 |
| | (0.426) | (0.128) | (0.131) | (0.118) | (0.111) | (0.106) |
| ORG_PAT_STOCK | 0.0979 | 0.138 | 0.161+ | 0.0296 | 0.0257 | 0.0352 |
| | (0.208) | (0.0891) | (0.0851) | (0.0789) | (0.0797) | (0.0688) |
| ORG_PRESTIGE | 0.256 | 0.0337 | | 0.0610 | 0.0419 | 0.0660 |
| | (0.192) | (0.0660) | | (0.0665) | (0.0629) | (0.0603) |
| TIMELAG | 0.455 | 0.665* | 0.639+ | 0.603* | 0.496* | 0.116 |
| | (0.341) | (0.319) | (0.326) | (0.255) | (0.248) | (0.512) |
| DIST0-20 | 1.265* | 1.547+ | 1.597+ | 0.986+ | 1.483* | -0.924 |
| | (0.581) | (0.857) | (0.874) | (0.547) | (0.675) | (0.711) |
| DIST20-50 | 0.235 | 0.310 | 0.380 | 0.218 | 0.205 | -0.655 |
| | (0.718) | (0.846) | (0.841) | (0.667) | (0.679) | (0.946) |
| DIST50-250 | 0.885 | 1.339* | 1.372* | 0.664 | 0.753 | -0.472 |
| | (0.592) | (0.617) | (0.615) | (0.565) | (0.622) | (0.387) |
| DIST250-1000 | 0.566 | 0.555 | 0.554 | 0.496 | 0.640+ | -1.114** |
| | (0.351) | (0.348) | (0.351) | (0.351) | (0.350) | (0.338) |
| DIST1000-2500 | 0.375 | 0.329 | 0.345 | 0.326 | 0.308 | -0.869* |
| | (0.348) | (0.353) | (0.356) | (0.352) | (0.339) | (0.351) |
| SAME_COUNTRY | -0.184 | -0.506 | -0.539 | -0.0913 | -0.142 | 1.733** |
| | (0.538) | (0.511) | (0.519) | (0.547) | (0.545) | (0.545) |
| PAPER_JIF^2 | -0.273 | | | | | |
| | (0.245) | | | | | |
| AUTHOR_PAT_STOCK^2 | -0.139 | | | | | |
| | (0.157) | | | | | |
| AUTHOR_PUB_STOCK^2 | -0.00181 | | | | | |
| | (0.0638) | | | | | |
| ORG_PRESTIGE^2 | -0.0383 | | | | | |
| | (0.0305) | | | | | |
| ORG_PAT_STOCK^2 | -0.00827 | | | | | |
| | (0.0331) | | | | | |
| CITY_PRESTIGE | | | -0.0158 | | | |
| | | | (0.0662) | | | |
| Observations | 1,638 | 1,234 | 1,234 | 1,638 | 1,638 | 1,071 |
| # articles | 313 | 264 | 264 | 313 | 313 | 378 |
| # twin-pat dyads | 771 | 580 | 580 | 771 | 771 | 500 |
| # simultaneous discoveries | 146 | 123 | 123 | 146 | 146 | 175 |
| Pseudo-R2 | 0.159 | 0.129 | 0.129 | 0.158 | 0.146 | 0.0943 |
| Log-likelihood | -482.1 | -376.1 | -376.4 | -483.1 | -489.7 | -339.2 |

Standard errors in parentheses; ** p<0.01, * p<0.05, + p<0.1

Notes: Dependent variable reports whether a paper-patent dyad is linked by an actual reference. OUTSIDE_R&D_HUBS refers to a paper published by an institution not within 50 miles of a "hub" of R&D activity for the patent subclasses associated with its simultaneous discovery. Standard errors are clustered at the level of the simultaneous discovery. Variables are defined in Table 1.

**Table 5:** The impact of the academic institution's location outside of an R&D hub: interactions

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Industry investment in institution | Industry investment in institution | Institutional prestige | Distance |
| OUTSIDE_R&D_HUBS | -1.449** | | | |
| | (0.488) | | | |
| PAPER_JIF | -0.752 | -0.662 | -0.352 | -0.273 |
| | (0.545) | (0.564) | (0.330) | (0.332) |
| PAPER_US | 0.414 | 0.0630 | 0.0978 | 0.336 |
| | (1.186) | (1.507) | (0.433) | (0.432) |
| PAPER_PATENTED | 0.102 | 0.173 | 0.457 | 0.326 |
| | (0.375) | (0.367) | (0.306) | (0.256) |
| AUTHOR_PAT_STOCK | 0.292 | 0.305 | 0.198 | 0.250 |
| | (0.263) | (0.270) | (0.199) | (0.192) |
| AUTHOR_PUB_STOCK | 0.0476 | -0.0107 | 0.0590 | 0.0139 |
| | (0.158) | (0.149) | (0.128) | (0.120) |
| ORG_PAT_STOCK | 0.0702 | 0.186 | 0.0912 | 0.0308 |
| | (0.110) | (0.137) | (0.0806) | (0.0783) |
| ORG_PRESTIGE | 0.0914 | 0.135 | | 0.0426 |
| | (0.104) | (0.0931) | | (0.0687) |
| TIMELAG | 1.578+ | 0.886 | 0.472 | 0.536* |
| | (0.828) | (0.886) | (0.352) | (0.272) |
| SAME_COUNTRY | -0.963 | -0.689 | -0.218 | -0.157 |
| | (1.111) | (1.458) | (0.545) | (0.562) |
| ORG_INDUSTRY_FUNDING | -0.0391* | | | |
| | (0.0182) | | | |
| OUTSIDEHUBS * ORG_INDUSTRY_FUNDING | 0.0375* | | | |
| | (0.0184) | | | |
| OUTSIDEHUBS & NO_FUNDING | | -2.367** | | |
| | | (0.801) | | |
| OUTSIDEHUBS & LOWER_FUNDING | | -1.277 | | |
| | | (1.395) | | |
| OUTSIDEHUBS & HIGHER_FUNDING | | -0.747 | | |
| | | (0.470) | | |
| OUTSIDEHUBS & HIGHEST_FUNDING | | -0.334 | | |
| | | (0.999) | | |
| OUTSIDEHUBS & LOWEST_PRESTIGE | | | -1.875** | |
| | | | (0.650) | |
| OUTSIDEHUBS & LOWER_PRESTIGE | | | -0.619 | |
| | | | (0.713) | |
| OUTSIDEHUBS & HIGHER_PRESTIGE | | | -0.986* | |
| | | | (0.438) | |
| OUTSIDEHUBS & HIGHEST_PRESTIGE | | | -0.582 | |
| | | | (0.400) | |
| OUTSIDEHUBS & DIST0-20 | | | | -0.297 |
| | | | | (1.114) |
| OUTSIDEHUBS & DIST20-50 | | | | 0.183 |
| | | | | (1.264) |
| OUTSIDEHUBS & DIST50-250 | | | | -1.925+ |
| | | | | (1.045) |
| OUTSIDEHUBS & DIST250-1000 | | | | -2.032** |
| | | | | (0.775) |
| OUTSIDEHUBS & DIST1000-2500 | | | | -1.612** |
| | | | | (0.605) |
| OUTSIDEHUBS & DIST>2500 | | | | -0.155 |
| | | | | (0.340) |
| Observations | 874 | 874 | 1,638 | 1,638 |
| # articles | 162 | 162 | 313 | 313 |
| # twin-pat dyads | 421 | 421 | 771 | 771 |
| # simultaneous discoveries | 78 | 78 | 146 | 146 |
| Pseudo-R2 | 0.204 | 0.242 | 0.164 | 0.173 |
| Log-likelihood | -242.8 | -231.1 | -479.4 | -474.3 |

Standard errors in parentheses; ** p<0.01, * p<0.05, + p<0.1

Notes: Dependent variable is whether a paper-patent dyad is linked by an actual reference. OUTSIDE_R&D_HUBS refers to a paper published by an institution not within 50 miles of a "hub" of R&D activity for its associated patent subclasses Standard errors clustered at the level of the simultaneous discovery. Base variables for the interactions of OUTSIDE_R&D_HUBS and dummies are omitted to save space.

**Figure 1: Construction of simultaneous-discovery dataset of dyads between each paper reporting the simultaneous discovery and each patent that referenced any of the papers reporting the simultaneous discovery. Each line represents a dyad in the dataset. Solid lines represent actual references while dotted lines represent unrealized references**.

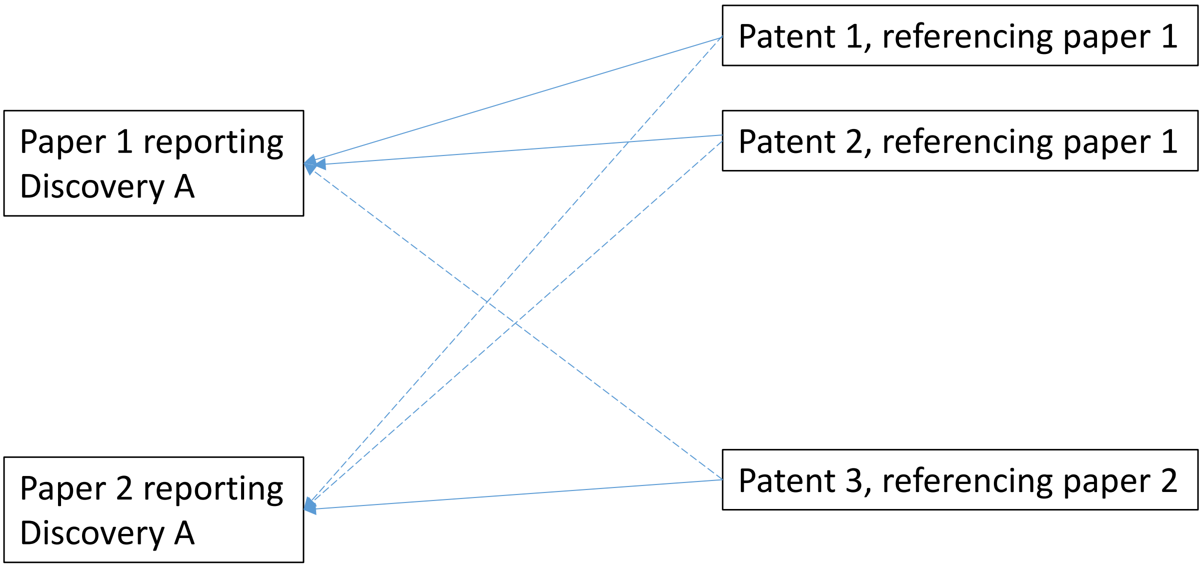
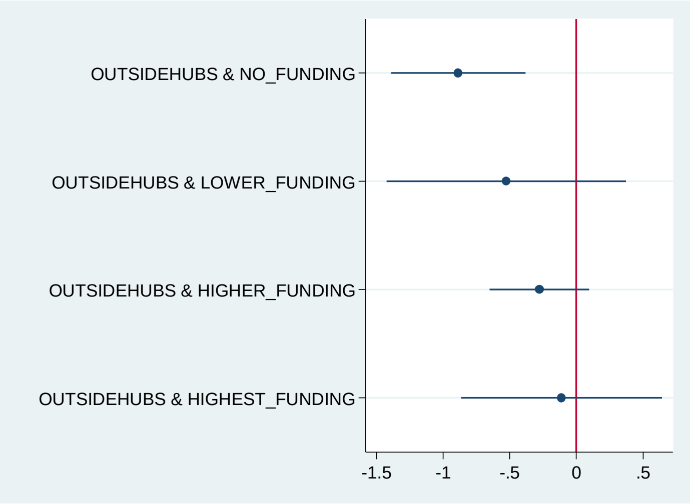
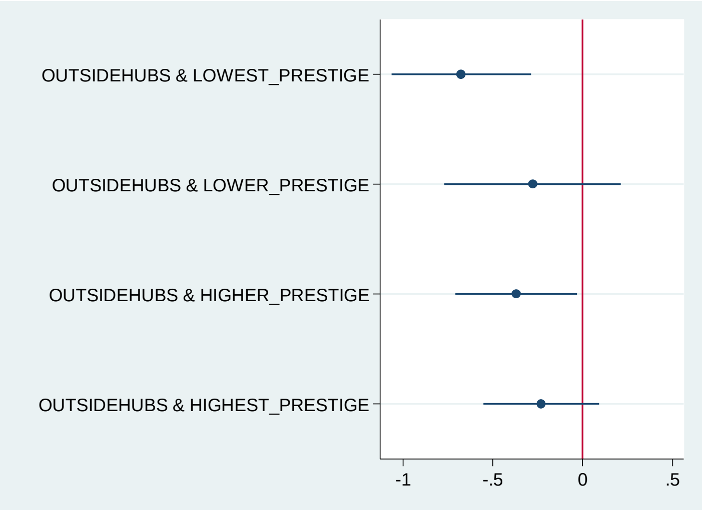


**Figure 2: Coefficient plots of interacting OUTSIDE_R&D_HUBS with other factors.**

**Panel A: Interaction of OUTSIDE_R&D_HUBS with various levels of industry funding**




**Panel B: Interaction of OUTSIDE_R&D_HUBS with various levels of organizational prestige**




**Panel C: Interaction of OUTSIDE_R&D_HUBS with various ranges of paper-to-patent distance**