

## *Internet adoption and knowledge diffusion* \*

**Chris Forman**

Scheller College of Business  
Georgia Institute of Technology  
800 West Peachtree St. NW  
Atlanta, GA 30308  
USA

[chris.forman@scheller.gatech.edu](mailto:chris.forman@scheller.gatech.edu)

**Nicolas van Zeebroeck**

iCITE  
Solvay Brussels School of Economics and Management  
Université libre de Bruxelles  
50, avenue F.D. Roosevelt - CP114  
1050 Brussels  
Belgium

[Nicolas.van.Zeebroeck@ulb.ac.be](mailto:Nicolas.van.Zeebroeck@ulb.ac.be)

*Complete Research*

### **Abstract**

*What is the capacity of ICT to reduce the (geographical and technological) localization of knowledge? In this paper, we analyze the impact of Internet adoption within US firms on knowledge spillovers. More specifically, we investigate the impact of basic Internet access on the likelihood that patents invented at a given R&D establishment cite patents invented elsewhere within the same firm. Our findings suggest that adoption of Internet significantly fosters cross-location citations in a significant way, and that these effects are proportional to the technological proximity of the establishments. This positive effect holds even when excluding collaborative patents or controlling for earlier collaborations, and suggests that Internet adoption has helped in reducing the spatial localization of knowledge but not in the ability to draw from new knowledge sources (i.e., across different technological areas).*

*Keywords: Geography of innovation, Internet adoption, IT investments, Knowledge spillovers, Patent citations, Technological proximity.*

---

\* We thank Harte Hanks Market Intelligence for supplying data. All opinions and errors are ours alone.

## 1 Introduction

Research and development operations are often physically dispersed, either across locations within a country (e.g., Leiponen and Helfat 2011, Miller, Fern, and Cardinal 2007), or across countries (Singh 2008, Penner-Hahn and Shaver 2005). These dispersed locations can be valuable, because they can provide the firm access to heterogenous labor pools, knowledge spillovers from public or private R&D, and adapt existing technological knowledge to new markets (e.g., Furman et al. 2005; Kogut and Zander 1993). However, even within firms, it is often difficult to effectively transfer and use knowledge produced elsewhere within the organization (Teece 1977). As a result, recent research has found somewhat mixed results about how multiple R&D locations influences the output and value of innovation (Leiponen and Helfat 2011, Singh 2008).

R&D organizations have a variety of means at their disposal to improve knowledge flows across locations. Prior researchers have found that knowledge flows will be facilitated in organizations with weak ties (Hansen 1999), that inventor movements can facilitate knowledge flows between the inventors past and present locations (Almeida and Kogut 1999; Singh and Agrawal 2011), and that the organizational design of research departments can also influence knowledge flows (Henderson and Cockburn 1994). It is widely believed that information and communication technologies (ICT), by lowering the costs of identifying useful information as well as the costs of transmitting it once it is found, also have the potential to increase the flows of valuable knowledge within firms (e.g., Alavi and Leidner 2001, Kleis et al. 2012). However, at the same time, research on knowledge management has argued that technologies that reduce the costs of knowledge circulation will not necessarily increase the flows of knowledge, particularly among members who had not already been communicating in the past (e.g., Brown and Duguid 1991, Vandenbosch and Ginzberg 1997). At present, due to data constraints, we have little direct systematic empirical evidence of the effects of ICT investments on knowledge flows within firms. This is particularly true for R&D knowledge flows, which as we describe later may have particularly unique challenges related to the transfer of complex or tacit knowledge. This is a significant gap in understanding. If ICT improves knowledge flows within firms, it has the potential to significantly reshape the traditional trade-offs associated with geographically dispersed R&D operations.

We take a first step toward providing systematic empirical evidence on these questions. Our approach uses the rapid declines in communication costs that occurred around the time of the initial commercialization of the Internet (e.g., Forman 2005; Forman and van Zeebroeck 2012). Our primary experiment examines changes in knowledge flows that occur between establishments within a firm after both establishments adopt Internet technology in the late 1990s. Specifically, we examine whether the adoption of basic Internet technology at two establishments increases the likelihood of a knowledge flow between them. When two firm locations adopt basic Internet technology, we find it increases the likelihood of a patent citation between them by 0.8 percentage points, or 13.3%. This number excludes citations made by patents resulting from collaborations between the two locations. We choose to focus on within-firm knowledge flows for two main reasons: the first is that it enables us to measure the direct outcome of a firm decision on its ability to reuse its own knowledge without relying on third-party actions. The second reflects in part a data constraint: patent citations are sometimes added for strategic motives, which is unlikely to affect differently different establishments within the same firm, whereas it might affect more strongly citation patterns to patents owned by third parties.

In a second step, we examine the influence of knowledge proximity between two establishments on the positive contribution of dyadic Internet adoption to knowledge flows. Based on Jaffe (1996)'s measure of proximity, we observe that knowledge flows increase as a result of dyadic Internet only when the two establishments are technologically close, i.e. when they share some knowledge base.

Our paper contributes to several fields of research. First, this paper contributes to the understanding of organizational practices that shape knowledge flows and knowledge exchange within geographically dispersed research organizations (e.g., Hansen 1999; Almeida and Kogut 1999; Henderson and Cockburn 1994; Zhao 2006). While this question has been explored from a variety of perspectives, to our knowledge there has been no systematic empirical study of the effect of ICT investments on

knowledge flows within organizations. The papers that are perhaps closest to ours are Kleis et al. (2012) and Forman and van Zeebroeck (2012). The former investigate the effects of ICT investments on patent output. The latter show, in a sample similar to this one, that investments in Internet technology are associated with an increase in geographically dispersed collaborations. However, that paper does not investigate the implications of IT investments on knowledge flows, as we do.

## **2 Theory and hypotheses**

### **2.1 The localization of knowledge and its roots**

Since the works of Mansfield and Teece, scholars have scrutinized the ability of firms and individuals to diffuse knowledge. In particular, Kogut and Zander (1992) have showed that knowledge does not easily cross firm boundaries and is easier to diffuse internally. Singh (2005) observed that intraregional and intra-firm knowledge flows are stronger than those across regional or firm boundaries. Singh and Marx (2013) highlighted that knowledge diffusion is also sensitive to political borders. Ever since, scholars have analysed what is now known as the stickiness or localization of knowledge, which refers to the difficulty in transferring or sharing knowledge at distance. With other words, knowledge tends to be disproportionately used in the same geography where it has been created.

Research on the organization of knowledge within firms has focused on a range of barriers to the identification and use of knowledge within organizations (see Alavi and Leidner (2001) and Foray (2004) for reviews). In this paper, we focus on two in particular that are likely to be influenced by the introduction of a new technology lowering communication costs: the costs of searching for and identifying relevant information and the costs of transferring information and using it once it has been identified.

One challenge is to identify the relevant knowledge within an organization. As has been noted elsewhere, organizations tend toward local search. Different arguments have been advanced to motivate this bias. The main argument raised generally refers to the strong path-dependence of R&D activities, which Cohen and Levinthal (1989) explain with the concept of absorptive capacity, itself forged by previous R&D efforts. But other arguments have also been raised, such as the resilience of organizational routines (Nelson and Winter, 1982), the strong contribution of accumulated knowledge stocks to the success of and returns to new R&D investments (Teece, 1988; Cohen and Levinthal, 1989), and the uncertainty surrounding the course of technological development and the evolution of economic and social environments (see e.g. Stuart and Podolny, 1996). According to Fleming (2001), uncertainty itself results from search and experimentation with new components and new combinations, themselves leading to greater variability in the outcome of research efforts.

A second problem is that even once relevant knowledge can be identified, it may be difficult to transfer. One question is why this might be the case in our setting, where we examine citations to patents. If knowledge is codified within patents, then the costs of knowledge transfer will be low. One might expect this to particularly be the case if there are strong incentives to disseminate scientific knowledge widely (Dasgupta and David 1994). However, important components of the knowledge embedded within patents may be tacitly held by engineers (Almeida and Kogut 1999, Agrawal 2006, Agrawal, Cockburn, and McHale 2006). For example, failed experiments may be important knowledge for understanding how to modify an invention for different applications but may not be written down because there is little incentive to do so (Agrawal, Cockburn, and McHale 2006).

### **2.2 Fostering knowledge acquisition and diffusion within firms**

Since the localization of knowledge has been established and sufficiently documented, the past decade of research in the field of innovation has produced a number of works focusing on factors that may help overcome this problem. Prominent contributions in this area include works on R&D collaborations (Cassiman and Veugelers, 2002) and foreign direct investments or international trade (Veugelers and Cassiman, 2004; MacGarvie, 2006). These results suggest that knowledge can overcome geo-

graphical and technological boundaries when supported by collaborations. For instance, in their study of large Japanese semiconductor producers, Stuart and Podolny (1996) show how Matsushita was the only firm in its industry to move away of local search, essentially by partnering with other firms giving them access to other technologies. The positive impact of collaborations on knowledge flows suggests that knowledge circulates well through personal interactions. It is therefore also likely to benefit from inventor mobility, as shown by Agrawal, Cockburn and McHale (2006), Azoulay, Graff Zivin and Sampat (2011) and Singh and Agrawal (2011).

The field of knowledge management (KM) is more generally interested in how organizations manage their intellectual capital in general and their technological knowledge in particular. Within larger knowledge management systems, IT has proved useful to provide organized knowledge repositories where an explicit representation of the firm's knowledge base is stored (Adamides and Karacapilidis, 2006) and to facilitate interaction and collaboration (e.g. Forman and van Zeebroeck, 2012), which themselves lead to the dissemination of knowledge (Adamides and Karacapilidis, 2006).

IT can reduce the costs of searching for and identifying knowledge within an organization by, for example, creating corporate directories. Intranets, knowledge repositories, and databases can provide exposure to a broad array of organizational information (Alavi and Leidner 2001). More generally, Internet-enabled communications, prominently email, can increase the frequency of interactions between distant workers. As shown in recent research (e.g. Reagans and McEvily, 2003), more frequent interactions between individuals fosters knowledge sharing between them. By strengthening the ties between inventors, Internet may therefore enhance their willingness and motivation to invest time, effort and energy in sharing knowledge. Once relevant information is located, IT facilitates knowledge transfer by lowering the costs of communication. In the case of transferring tacit knowledge, IT may reduce the costs of communicating with holders of relevant knowledge. In other words, once relevant knowledge is identified, in cases where it cannot be easily codified then IT reduces the costs of person-to-person knowledge transfer (Hansen, Nohria, and Tierny 1999). However, there is one caveat. While IT is effective at lowering the communication costs among existing ties, it is less effective at facilitating the formation of new ties (Gaspar and Glaeser 1998). This is particularly true for the margin of IT that we consider, where most IT facilitated bilateral communication, in contrast with tools to expand social networks and to use IT to develop new ties (McAfee 2009). Thus, in the absence of these connections, IT may have little impact on knowledge flows.

In this paper, we investigate the contribution of IT adoption on knowledge dissemination within firms. Our focus is on a very specific margin of IT investments, which is access to the early years of the commercial Internet. This includes basic services such as WWW services (including Intranets) and SMTP-based email. This focus does therefore not include more sophisticated systems to support collaborative work or knowledge repositories. Rather, the main interest of the focal technology resides in a decrease in communication costs at distance. We therefore expect lower communication costs to result in additional interactions between distant researchers and hence technological information sharing between them, a prediction that is summarized in Hypothesis 1.

---

**Hypothesis 1:** Adoption of basic Internet will be associated with an increase in the likelihood of knowledge flows between geographically dispersed teams.

---

We next investigate under what conditions IT will be most effective at facilitating knowledge transfers. We first consider variations in technological distance. If knowledge is tacit or depends on the absorptive capacity of the recipient, then transfer of the knowledge will depend upon person-to-person knowledge transfer. The KM literature offers an alternative view on this prediction: if the explicit knowledge that is shared across a computer-network requires complementary tacit knowledge, the explicit knowledge transmitted will not be usable by its addressee as such. This will likely be the case when the knowledge base of the destination is different from that of the source. Information technolo-

gy can reduce the costs of this type of person-to-person knowledge transfer. However, this type of transfer depends upon locating the right person within the organization and the source showing the willingness to transfer the knowledge. If the knowledge is technologically related to that of the recipient, she may be able to use existing strong ties with inventors in the same or related groups to transfer it (Granovetter 1973). As has been noted elsewhere, however, the transfer of more distant knowledge may depend on the use of weak ties within the organization (Granovetter 1973, Burt 1995, Hansen 1999). The margin of IT that we investigate will be less effective at forming new ties. In fact, early theoretical work has suggested that this type of IT may cause users to communicate more with sources with similar backgrounds (McAfee 2009). To recap, searchers for knowledge in related technological areas will find it easier to convert awareness of information about a technology into useful knowledge that can be incorporated into an invention. When some of the knowledge about how to use the technology is tacit, they will be more likely to have strong ties with inventors that can assist them in how to implement the invention. IT will be effective at lowering the search costs of identifying the relevant technologies, and will also help facilitate communication with strong ties.

Searchers for knowledge in more distant technology areas will find it more difficult to use IT to lower the costs of knowledge transfer. While IT may help inventors find information about more distant technologies, they will be unable to convert information about the technology into useful knowledge without person-to-person contact with inventors. In other words, we expect the Internet to be inefficient at diffusing technological information across technologically distant teams, which is summarized in Hypothesis 2. We therefore expect Internet adoption to be associated with higher knowledge spillovers across geographical distance but not across technological distance.

---

**Hypothesis 2:** Adoption of basic Internet will be associated with a smaller increase in the likelihood of knowledge flows among technologically distant teams.

---

### 3 Empirical strategy

We argue that adoption of basic Internet will be associated with a decline in the costs of knowledge transfer between two establishments. As a result, we expect an increase in knowledge flows between two firm establishments adopting basic Internet. Specifically, our empirical strategy consists in estimating the incidence of knowledge flows between any two geographically-distant establishments within a given firm, using the adoption of Internet by both locations as a treatment. We use fixed effects panel regressions to identify the treatment effect, in a setting where time represents two-year periods. The choice of two-year periods reflects a data constraint (we are able to observe Internet adoption only in even years) but this choice also allows for enough time for Internet and its effects to diffuse within establishments and firms. The unit of observation in our data is a firm dyad-year, and our use of fixed effects will difference out the effects of time-invariant dyad unobservables that may increase the incidence of citations between the two establishments. Our approach yields the following estimating equation:

$$Citation_{ijkt} = \alpha_1 X_{ijkt} + \alpha_2 Z_{ijkt} + \beta Internet_{ijkt} + \mu_{ijk} + \tau_t + \varepsilon_{ijkt} \quad (1)$$

$Citation(ijkt)$  indicates the existence of at least one US patent of firm  $i$  that was applied for in year  $t$  or the preceding year, that was invented in location  $j$ , and that cites another patent invented in location  $k$  of the same firm in the 5 preceding years (following Hall et al. (2001) and many others, we deal with data truncation issues affecting patent citations by accounting only for 5-year forward citations).  $x(ijkt)$  is a vector of time-varying controls at the establishment-pair level,  $z(ijkt)$  is a vector of time-varying controls for local characteristics. Our main variable of interest is  $Internet(ijkt)$ , which indicates whether establishments  $j$  and  $k$  of firm  $i$  had both adopted Internet in year  $t$ .  $\mu(ijk)$  measures pair fixed-effects and  $\tau(t)$  measures time fixed-effects. Hypothesis 1 will be verified if  $\beta$  is found significantly larger than 0. For all of our regressions we use robust standard errors clustered at the firm-dyad level.

In a second step, we explore the sensitivity of our results to the exclusion of collaborative patents. Specifically, a positive coefficient on  $\beta$  may reflect increases in collaborations between locations  $j$  and  $k$  that are themselves precipitated by the adoption of Internet; for example, Forman and van Zeebroeck (2012) find that when two firm locations adopted Internet over this period the likelihood of a research collaboration between them increased by 23.0%. To address this potential concern, we identify patents that were co-invented in the two locations and exclude them from our measures of citations.

We utilize the fact that in the first two years of our data, 1992 and 1994, adoption of basic Internet within firms in our sample will be equal to zero. This is because these two years predate the commercialization of the Internet. For the parameters in equation (1) to be identified, we require significant within-firm variance in basic Internet adoption within firms in 1996 and 1998, over the period when the commercial Internet began to diffuse. Using a similar sample, Forman and van Zeebroeck (2012) demonstrate significant variance across locations within firms; in general, firms did not adopt basic Internet across all establishments at the same time. As they note, there are likely to be several reasons for this variance. First, there was significant geographic variance in the cost of business Internet adoption over this period (Forman, Goldfarb, and Greenstein 2005). Further, firm locations differed in their legacy IT investments in ways that could shape the costs to adoption (Forman 2005). Last, because governance of the IT function is frequency decentralized, (Sambamurthy and Zmud 1999; McElheran 2014), firm locations may maximize local rather than firm-wide benefits in adoption decisions.

We run several tests to probe the validity of our estimates. First, we control for various sources of heterogeneity at the firm, establishment, location and pairs levels. Second, we conduct several falsification tests to show that the effects of Internet adoption show up only where they are expected, given our theoretical framework. And finally, we also show that our results are robust to the use of instrumental variables. Our instruments proxy for geographic variance in the costs to adoption across locations.

In order to test our second hypothesis we compute a measure of technological proximity between the two sides of each pair and estimate its joint effect with Internet, yielding the following equation:

$$Citation_{ijkt} = \alpha_1 X_{ijkt} + \alpha_2 Z_{ijkt} + \beta Internet_{ijkt} + \gamma Internet_{ijkt} \times Proximity_{ijkt} + \mu_{ijk} + \tau_t + \varepsilon_{ijkt} \quad (2)$$

Where  $Proximity(ijkt)$  measures the technological proximity of locations  $j$  and  $k$  of firm  $i$  at time  $t$ . In this case, Hypothesis 2 will be validated if  $\beta$  is significantly different from 0 and  $\gamma$  is found significantly larger than 0. Both equations are estimated using a fixed effects linear probability model.

## 4 Data

Our data come from a variety of sources. We match data on IT investment from a well-known private data source to data on patent citations from the USPTO. We combine these data with information from Compustat (to obtain controls related to R&D and firm size) and from the U.S. Census County Business Patterns data (to obtain data for regional controls).

### 4.1 Patent Data

Within each firm, we then use citations between patents invented at different locations as a proxy for within-firm across-establishments knowledge spillovers. To do so, we use data on patents filed by multi-establishment US manufacturing firms at the USPTO. We use the application date as the date for the citing patent because of delays in the application-to-grant period, and because application dates are closer to when the invention occurred (e.g., Griliches 1990). Our key variable is equal to whether there was a citation from patents with application date  $t$  from location  $j$  to location  $k$  over the previous five year period. Our focus on the extensive margin of whether there exist any citations in part reflects the distribution of our dependent variable; the mean number of citations across our data set in any one year is 0.31 (Table 1), however only 5.8% of dyads have a citation between them. However, we experiment with using the number of citations as our dependent variable within a series of Poisson count data regressions, and our results are robust. While we acknowledge the limitations of using patent cita-

tions as a proxy for knowledge flows (e.g., Alcacer and Gittelman 2006, Lampe 2012, Roach and Cohen 2013)—in particular, not all citations reflect knowledge flows and some knowledge flows will not be reflected in citations—our focus on self-citations will mitigate some of these concerns, as organizations will have incentives to include appropriate citations to increase the appropriability from their inventions (Hall et al. 2005).

Our analysis requires us to identify both the firm and location in which a patent is invented. We map patents to firms using the assignee field from the patent and the GVKEY of the COMPUSTAT database using the matching files from the National Bureau of Economic Research (NBER) Patent Data Project. Using this procedure, we obtained the universe of patents with a matching GVKEY that were applied for during the period 1990-1998. As noted above, the unit of analysis in our data will be a firm location pair. We aggregate our data to firm-MSAs, rather than study particular addresses of plants. This reflects a data constraint; the USPTO patent data list only the city and state of an inventor, and so we are unable to identify the particular establishment that an inventor works at within an MSA. Using the city and state of the inventor listed in the patent, we map this information to zip codes and then in turn match zip codes to MSAs. When consolidated MSAs (CMSAs) were present, we used those because they better captured commutation patterns. In regions where inventors resided outside of MSAs, we constructed “phantom MSAs” which consisted of the areas of a state outside of all of the MSAs. MSAs are widely used as a unit of analysis in studying the geography of innovation (e.g., Feldman and Audretsch 1999), but our procedure may assign some patents to the incorrect MSA when one inventor commutes to or from a different MSA.

## **4.2 Information Technology Data**

Our data on Internet adoption comes from a private market intelligence source, the Harte Hanks Market Intelligence Computer Intelligence Technology database (hereafter, CI database). The database contains a wide range of information related to establishment- and firm-level data on investments in information technology hardware, software, and networking, as well as data related to demographic information such as the number of employees and industry of the establishment and firm. The data have been used in a wide range of studies related to the adoption of IT (Bresnahan and Greenstein 1996, Forman, Goldfarb, and Greenstein 2005) and IT productivity (Brynjolfsson and Hitt 2003; Bloom, Sadun, and Van Reenen 2012). More closely related to our setting, it has been used to study the role of ICT investments in reducing the costs associated with economic and geographic isolation in other settings (e.g. Forman 2005, Forman, Goldfarb, and Greenstein 2005) and on the effects of IT investments on the organization of R&D (Forman and van Zeebroeck 2012).

As noted in prior work (e.g., Forman 2005; Forman, Goldfarb, and Greenstein 2005), the CI database contains a wide range of information related to an establishment’s adoption of IT. For this paper, our interest is in exploring the implications of a margin of Internet that lowers communications costs across establishments, but which imposes little direct change on the business processes of organizations. The set of Internet technologies that we study will require little adaptation by organizations, and will involve many of the technologies that diffused around the initial commercialization of the Internet. Our interest in this particular set of technologies reflects both the time period we study (around the initial commercialization of the Internet) as well as our interest in exploring how a set of technologies that reduce the costs of basic communications will change the nature of knowledge flows within firms.

We consider an establishment to have adopted basic Internet when any one of the following occurs: the establishment reports that it has an Internet Service Provider (ISP); the establishment reports having an internal intranet based on the TCP/IP protocol (Transmission Control Protocol/Internet Protocol); the establishment reports using the Internet for research purposes. We consider that no establishments have adopted Internet in 1992-1994 as this period predates the launch of commercial Internet. Among the 37,720 pairs of locations, 0% had adopted basic Internet in 1992 and 1994, 12% had adopted by 1996, and 58% had adopted by 1998. As noted above, the CI data are collected at the establishment level. To match our establishment-level IT data to the patent data described above, we

match establishments to MSAs. We first match the firm identifier in the CI database to a COMPUSTAT GVKEY, and match establishments to MSAs using the establishment zip code. Whenever we have several establishments within a given MSA, we take the average of all variables, except for the adoption of Internet where we consider that a location has adopted as soon as one establishment within the location has.

### **4.3 Firm-MSA Pairs**

The focus of our study is to examine the effects of IT investments on cross-establishment knowledge flows within organizations. We estimation equation (1), which allows us to measure whether adoption of basic Internet in firm locations  $j$  and  $k$  in year  $t$  is associated with a citation from location  $j$  to location  $k$  and vice-versa. To do this, we form the complete set of potential firm-location pairs within an organization, and examine whether there exists a patent in location  $k$  that cites a patent  $j$  over the five-year period preceding time  $t$  (and vice-versa). The dataset is symmetric, which means that we keep both combinations of every set of two locations (A-B and B-A). The dependent variable is indeed different in the two configurations: it will reflect citations by patents invented in location A to patents invented in location B in the first case, and citations from B-patents to A-patents in the second. We restrict our estimation sample to firm-MSA-year triplets where the firm is in the manufacturing industry (Standard Industrial Classifications 20-40) and to firm-MSAs in which there is at least one patent in two separate years during the period 1992-1998. These conditions are to restrict our sample to firm establishments engaged in research activities.

### **4.4 Controls**

We control for a variety of firm- and location-specific factors in our regressions. To control for variance in R&D inputs across firms, we compute the flow of R&D spending (in dollars) using COMPUSTAT and normalize this figure by dividing total spending by the number of firm-locations in our data. We use the Harte Hanks data to compute firm-location employment as the sum of establishment employment across establishments in the location. We compute the log of the average employment across the two locations to estimate equation (1).

To control for how technological similarity between two establishments influences the likelihood of observing a citation, we compute technological proximity based on Jaffe (1986) and MacGarvie (2006), which consists in computing the share of patent portfolios that fall in the same technological classes. Following Benner and Waldfoegel (2008), we consider all USPC classes assigned to patents in our sample in order to minimize biases in our measure. Our results are however robust to the use of the main technological class only. For some firm-location pairs this variable was undefined because one of the establishments in the pair had no patent in the period considered. In this case we added a dummy variable to indicate that proximity is undefined. We further control for other potential sources of heterogeneity such as patent stocks at the establishment and county levels. Additional county-level controls include the share of local employment in manufacturing, local average weekly wages, and the log of local employment. Our estimates further include location-pair fixed effects, which further reduces the potential impact of other sources of unobserved heterogeneity such as prior investments in alternative technologies like databases or knowledge repositories.

### **4.5 Descriptive statistics**

Descriptive statistics for all variables in the model are presented in Table 0 in the appendix. Among the 37,720 pairs of locations, 0% had adopted basic Internet in 1992 and 1994, 12% had adopted by 1996, and 58% had adopted by 1998. Overall, 6% of the pairs are linked by at least one directional citation (the average number of citations within a pair is 0.31 with a maximum of 372).

## 5 Empirical results

### 5.1 Baseline results

We begin with our econometric estimation of Equation 1 to explore the impact of Internet adoption on pairwise citations (Hypothesis 1). The baseline results are reported in Table 1. As is common in linear probability models (e.g., Athey and Stern 2002), the overall R<sup>2</sup> (computed by excluding the fixed effects in the R<sup>2</sup> computation) is low. The first column shows the results of our main estimation with our main set of controls. The coefficient associated with Internet adoption is positive and statistically significant (at the 1% level), suggesting that Internet does foster knowledge flows between distant establishments within firms: if both locations adopt Internet, this translates into a 1.1 percentage point increase in the likelihood of citations between them. Given that the mean likelihood of within-pair citation is 5.8% in our sample, this implies that Internet adoption increases the incidence of citations by 19.0%. The incidence of citations is also strongly affected by the technological proximity between the two locations forming a pair. Columns 2 and 3 of Table 1 report results of alternative specifications. In column 2, we include an extended set of controls, adding 10-years and current stocks of patents accumulated within the pair. In column 3, we explore the robustness of our results to the use of an unbalanced panel. Our results are robust to all of these.

As noted above, the results in the first three columns of Table 1 provide an upper bound for the effects of the Internet on knowledge flows; some of the results will reflect the effects of the Internet on promoting knowledge flows through collaborations, rather than the effects on knowledge flows per se. Columns 4 and 5 of Table 1 show the robustness of our results to a citation measure that excludes citations from patents resulting from collaborations between inventors located in the two locations. These results will provide a lower bound for the effects of the Internet on citations. Although the point estimate of the treatment effect is slightly smaller (Internet adoption is associated with a 0.8 percentage point increase in the likelihood of citation flows v. a 1.1 percentage point increase with collaborative patents included), our core result remains economically and statistically significant (at the 5% level). As this citations measure is the more conservative one, we use it in all of our remaining regressions.

### 5.2 Exploring Robustness

Before exploring how technological proximity mediate our core result, we do a number of tests to assess the validity of our core assumption, which is that no unobserved factors that are correlated to Internet adoption and patent citations affect our results. We first provide the results of a series of falsification exercises, presented in Table 2. Column 1 reports the result of an alternative specification that includes a dummy for Internet when it is adopted only at one of the two locations in the pair. The results indicate that the effect of Internet materializes only when both establishments adopt. This is supportive of our main argument that Internet adoption reduces the costs of sharing technological information at distance and cut against the alternative view that Internet gives access to new repositories of information such as online databases. Column 2 investigates whether we observe the right timing in the effect of Internet adoption on knowledge flows. To do so, we include measures that are turned on two years in advance of the true adoption date. If the benefits of Internet adoption show up prior to when it is actually adopted, this would raise concerns that our results may reflect the effects of unobserved factors. Column 2 shows that a measure of Internet adoption 2 years in advance of the adoption date has no effect on citations. The lack of appearance of the benefits of Internet appearing before it is actually adopted raises confidence in our findings. Column 3 explores whether the benefits of Internet adoption appear with a lag; in other words, whether it takes some time for new citation patterns to emerge after the initial adoption of Internet technology. Prior work that has explored the productivity implications of IT investment has showed that the benefits of IT often appear with a lag (e.g., Brynjolfsson and Hitt 2003). We find that lagged Internet adoption has no incremental effect on the propensity to cite. This could reflect the fact that the margin of Internet we study was relatively

straightforward to implement and so its benefits appear immediately, or alternatively the lack of power in our statistical test.

To further address concerns about omitted variable bias, in Table 4 we present the results of instrumental variables estimates. We explore the implications of three instruments, each of which will influence the costs of adopting basic Internet. Our first two instruments capture changes to regulatory policy. They are the year in which the state in which the location in the pair has adopted rate of return regulation, and the year in which the state has instituted a price cap or freeze on telecommunications services. As we do for our other variables, we take the average across the two locations in the pair. These two variables can influence the likelihood of basic Internet adoption in two ways, in potentially opposite directions. First, by directly lowering the costs of purchasing telecommunications services, they may directly influence the costs of adoption. Second, as Greenstein and Mazzeo (2006) note, this variable can capture local variance in regulatory stringency. For example, states that have adopted rate of return regulation later may have a more welcoming attitude toward experimenting with competition, which may translate into lower costs for a competing competitive local exchange carrier. This friendlier attitude toward competition may translate into increased entry and so lower costs of procuring telecommunications services and so lower adoption costs.

We also instrument using the number of ARPANET nodes in the MSA. The ARPANET was a wide area network that was a predecessor to the Internet. Increases in this variable will represent increasing local familiarity with Internet technologies. Forman, Goldfarb, and Greenstein (2005, 2008) argue that such local capabilities and expertise can lower the costs of adopting Internet technologies. Further, because the number of ARPANET nodes represent historical decisions by the Department of Defense of U.S. university networks, they are unlikely to be correlated with Internet adoption decisions during our sample period. As noted earlier, Internet adoption is zero in 1992 and 1994. As a result, we interact each instrument with a dummy variable that is turned on during 1996 and 1998 and zero otherwise.

Table 4 presents the second-stage results. While the sign of the second-stage coefficient on basic Internet remains stable across specifications, there are some differences in both the size and statistical significance of the instrument. As noted above, our instruments have additional power when included together. In general, the second stage coefficient estimates are larger than those in the baseline estimates without the instruments. For example, the coefficient on basic Internet in column 4 of Table 4 is 0.1319, while the coefficient is equal to 0.0080 in our baseline estimates for the (excluding collaborative citing patents) in column 1 of Table 1. There could be several reasons for this pattern. First, our basic Internet variable may be measured with error, and our instruments may help to remove some of this measurement error (e.g., Hausman 2001). Second, our results may reflect a local average treatment effect. While our results may be valid in the sense that they are uncorrelated with citation patterns but for their impact on basic Internet adoption, it may be that firm location pairs whose Internet adoption behavior is influenced by variance in the instruments will also have a particularly large increase in citations resulting from basic Internet adoption (Angrist, Imbens, and Rubin 1996). In sum, our instrumental variables estimates provide additional support for a causal interpretation of our results. While on their own they may be unable to provide convincing evidence in support of such an interpretation, in conjunction with our other falsification analyses and additional tests they form part of a larger picture that Internet was correlated with an increase in cross-establishment citation.

### 5.3 Exploring the Effects of Technological Proximity

In this section we study whether the effects of adopting basic Internet on citation patterns are greater when the citing and cited locations are technologically proximate or have collaborated in the past. Our measure of proximity is based upon Jaffe (1986) and MacGarvie (2006) as described above in section 4.4. We use both a continuous measure of proximity as well as a dummy variable that indicates the pair is in the top quartile of proximity, for ease of interpretation. Columns 1 and 3 of Table 5 show that the effects of basic Internet on citations only appear when the citing and cited locations are technologically close. In particular, column 3 shows that locations that adopt basic Internet in locations that are

proximate increase the likelihood of a citation by a statistically and economically significant 3.1 percentage points. In contrast, locations that adopt basic Internet but are not technologically proximate see no growth in citations. Columns 2 and 4 provide the instrumented results. To construct the instruments, we use the same three instruments as in column 4 of table 4 and interact them with our proximity measures. Thus, we have six instruments in total: the three instruments in table 4 along with their interactions with proximity. The results are directionally similar to those without instruments, and the coefficients of both basic Internet and its interaction with proximity are larger in magnitude (more positive than those in columns 1 and 3).

## **6 Discussion and implications**

Our results are consistent with earlier works in innovation and knowledge management. The absence of impact of Internet on cross-technology flows may be supported by different interpretations: the resilience of strong organizational routines (Nelson and Winter, 1982), the uncertainty associated with search and exploration combining knowledge from different fields (“recombinant uncertainty” as coined by Fleming (2001)), and the returns to accumulated knowledge stocks (Teece, 1988; Cohen & Leinthal, 1990). The theories supported in these works point at the higher risks and costs associated with recombinant research, which act as disincentives for firms to engage in such activities. Our results suggest that basic Internet may not be the right instrument to overcome these.

In the knowledge management literature, a difference is often made between tacit and explicit knowledge as complements of one another. By nature, the type of technology we focus on (basic Internet access, which means email or static web pages) can only help transfer knowledge that has been made explicit and digitized, leaving the tacit knowledge unmoved. McDermott (1999) argues that unlike information, knowledge requires personal experimentation and context to be acquired and can therefore not be simply transmitted. Within their own field of expertise, research teams have accumulated knowledge and therefore developed their own contextual referential that will enable them to absorb additional explicit knowledge without the corresponding background and experimentation. This directly refers to the notion of absorptive capacity. But it seems reasonable to assume that research teams are lacking the required background or contextual knowledge to acquire and absorb crude explicit knowledge from outside their own field of expertise.

These findings speak to the blooming literature on the contribution of ICT to the growth in intangibles (e.g., Gao and Hitt 2004, Kleis et al. 2012) by uncovering one concrete mechanism through which such contribution takes place, namely a better circulation of technological knowledge at distance between researchers active in the same field. But they also illustrate some of the potential limits to the returns on and value of ICT investments, in this case the need for other types of investments (e.g. human or organizational) to diffuse and recombine knowledge across different technological fields. Our results also have implications for the organization of R&D activities, as they suggest that ICT may reduce to some extent the geographical concentration of technological knowledge and facilitate the sharing of information between distant knowledge workers. This finding has a potential to affect the trade-off between centralized and decentralized organizations of R&D activities.

Our conclusions are however limited in at least two dimensions. The first is whether our results would extend to more advanced technologies such as social networks and collaborative tools. The second is that we look only at intra-firm knowledge flows, whereas a broader question is whether ICT investments may also help cross institutional borders. We hope our results will stimulate further works.

	Baseline	Extended controls	Unbalanced panel	Baseline (excl. collaborations)	Ext. controls (excl. collaborations)
Basic Internet in both locations	0.0111*** (0.0041)	0.0102** (0.0040)	0.0095** (0.0040)	0.0080** (0.0038)	0.0072* (0.0038)
Log of per-establishment R&D spending	0.0170*** (0.0041)	0.0022 (0.0040)	0.0206*** (0.0036)	0.0144*** (0.0037)	0.0028 (0.0036)
Proximity in research among two locations	0.0730*** (0.0108)	0.0594*** (0.0105)	0.0682*** (0.0095)	0.0608*** (0.0100)	0.0503*** (0.0098)
Proximity variable not applicable	-0.0053*** (0.0018)	0.0149*** (0.0022)	-0.0044*** (0.0016)	-0.0058*** (0.0016)	0.0099*** (0.0019)
Log of establishment employees	-0.0109 (0.0090)	-0.0116 (0.0089)	-0.0030 (0.0088)	-0.0099 (0.0082)	-0.0104 (0.0082)
Share of local employment in manufacturing	0.1784 (0.1771)	0.1622 (0.1753)	0.2115 (0.1555)	0.1778 (0.1629)	0.1647 (0.1617)
Local average weekly wages	0.0002** (0.0001)	0.0002* (0.0001)	0.0002** (0.0001)	0.0002** (0.0001)	0.0001 (0.0001)
Log of local employment	-0.0216 (0.0431)	-0.0662 (0.0430)	-0.0467 (0.0390)	-0.0178 (0.0399)	-0.0525 (0.0399)
Log of number of local patents	0.0016 (0.0117)	-0.0070 (0.0116)	0.0232** (0.0103)	0.0079 (0.0111)	0.0013 (0.0110)
Log of patent stock over previous 10 years		0.0326*** (0.0034)			0.0251*** (0.0030)
Log of patent stock in current period	0.2070 (0.5671)	0.0300*** (0.0018)	0.3377 (0.5113)	0.1286 (0.5254)	0.0234*** (0.0016)
Constant	0.01	0.02	0.01	0.01	0.02
R <sup>2</sup>					
N	37,720	37,720	51,340	37,720	37,720

Table 1. Baseline results (including and excluding citing collaborative patents). Robust standard errors, clustered at the firm-location pair. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Year and Pair fixed effects included but not reported

	Any Internet adoption	Internet adoption leads	Internet adoption lags
Basic Internet in both locations	0.0092** (0.0041)	0.0104** (0.0049)	0.0095** (0.0041)
Basic Internet at any of the two locations	0.0017 (0.0039)		
Adopt Internet 2 years in the future		0.0011 (0.0035)	
Lagged Internet adoption			-0.0024 (0.0075)
Log of per-establishment R&D spending	0.0174*** (0.0039)	0.0174*** (0.0039)	0.0174*** (0.0039)
Proximity in research among two locations	0.0678*** (0.0107)	0.0678*** (0.0107)	0.0678*** (0.0107)
Proximity variable not applicable	-0.0051*** (0.0017)	-0.0051*** (0.0017)	-0.0051*** (0.0017)
Log of establishment employees	-0.0133 (0.0086)	-0.0133 (0.0086)	-0.0134 (0.0085)
Share of local employment in manufacturing	0.3512** (0.1711)	0.3503** (0.1710)	0.3499** (0.1710)
Local average weekly wages	0.0002** (0.0001)	0.0002** (0.0001)	0.0002** (0.0001)
Log of local employment	-0.0585 (0.0419)	-0.0588 (0.0419)	-0.0588 (0.0418)
Log of number of local patents	0.0100 (0.0111)	0.0100 (0.0111)	0.0101 (0.0111)
Constant	0.6459 (0.5525)	0.6494 (0.5524)	0.6505 (0.5519)
R <sup>2</sup>	0.01	0.01	0.01
N	37,720	37,720	37,720

Table 2. Robustness estimates. Robust standard errors, clustered on firm-location pair, in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Year and Pair fixed effects included but not reported.

	First change to ROR	Number of ARPANET Nodes	First price cap or freeze	All instruments	LIML
Basic Internet in both locations	0.0982 (0.0632)	0.3638 (0.2628)	0.0782 (0.1343)	0.1319** (0.0527)	0.1335** (0.0535)
Log of per-establishment R&D spending	0.0126*** (0.0043)	0.0074 (0.0072)	0.0130*** (0.0049)	0.0119*** (0.0043)	0.0119*** (0.0043)
Proximity in research among two locations	0.0601*** (0.0107)	0.0581*** (0.0125)	0.0603*** (0.0106)	0.0599*** (0.0108)	0.0599*** (0.0108)
Proximity variable not applicable	-0.0049*** (0.0018)	-0.0021 (0.0038)	-0.0051** (0.0021)	-0.0045** (0.0018)	-0.0045** (0.0019)
Log of establishment employees	-0.0130 (0.0085)	-0.0223 (0.0140)	-0.0123 (0.0095)	-0.0142* (0.0086)	-0.0143* (0.0086)
Share of local employment in manufacturing	0.2161 (0.1765)	0.3289 (0.2657)	0.2076 (0.1788)	0.2304 (0.1795)	0.2311 (0.1798)
Local average weekly wages	0.0002** (0.0001)	0.0002 (0.0001)	0.0002** (0.0001)	0.0002** (0.0001)	0.0002** (0.0001)
Log of local employment	-0.0094 (0.0438)	0.0153 (0.0590)	-0.0113 (0.0476)	-0.0063 (0.0449)	-0.0062 (0.0450)
Log of number of local patents	0.0033 (0.0119)	-0.0101 (0.0183)	0.0044 (0.0142)	0.0016 (0.0120)	0.0016 (0.0120)
R <sup>2</sup>	-0.02	-0.35	-0.01	-0.04	-0.04
N	37,720	37,720	37,720	37,720	37,720

Table 3. Instrumental variables estimates, second stage. Robust standard errors, clustered on firm-location pair, in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Year and Pair fixed effects included but not reported.

	Baseline, Fixed proximity, continuous measure	IV, Fixed proximity, continuous measure	Baseline, Fixed proximity, top quartile	IV, Fixed proximity, top quartile
Basic Internet in both locations	-0.0018 (0.0037)	0.1304** (0.0573)	0.0006 (0.0036)	0.1637*** (0.0591)
Basic Internet x Fixed proximity (continuous measure)	0.0571*** (0.0179)	0.1542*** (0.0336)		
Basic Internet x Fixed Proximity (upper quartile measure)			0.0305*** (0.0083)	0.0644*** (0.0154)
Proximity in research among two locations		0.0592*** (0.0112)		0.0589*** (0.0112)
Proximity variable not applicable		-0.0036* (0.0019)		-0.0034* (0.0020)
Log of per-establishment R&D spending	0.0162*** (0.0040)	0.0091** (0.0044)	0.0164*** (0.0040)	0.0096** (0.0044)
Log of establishment employees	-0.0093 (0.0080)	-0.0109 (0.0088)	-0.0093 (0.0080)	-0.0126 (0.0090)
Share of local employment in manufacturing	0.1869 (0.1691)	0.2903 (0.1838)	0.1940 (0.1691)	0.3043 (0.1888)
Local average weekly wages	0.0002** (0.0001)	0.0001 (0.0001)	0.0002* (0.0001)	0.0001 (0.0001)
Log of local employment	-0.0159 (0.0426)	0.0018 (0.0456)	-0.0170 (0.0426)	0.0002 (0.0461)
Log of number of local patents	0.0102 (0.0116)	-0.0001 (0.0122)	0.0101 (0.0116)	-0.0013 (0.0122)
Constant	0.0863 (0.5572)		0.1009 (0.5577)	
R <sup>2</sup>	0.00	-0.06	0.01	-0.08
N	37,720	37,720	37,720	37,720

Table 4. IV Results for proximity interactions. Robust standard errors, clustered on firm-location pair, in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Year and Pair fixed effects included but not reported.

## References

- Adamides, E. and N. Karacapilidis (2006), Information technology support for the knowledge and social processes of innovation management, *Technovation* 26, 50-59.
- Agrawal, A. 2006. Engaging the inventor: exploring licensing strategies for university inventions and the role of latent knowledge. *Strategic Management Journal* 27: 63-79.
- Agrawal, A., I. Cockburn, and J. McHale. 2006. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography* 6: 571-591.
- Agrawal, A. and A. Goldfarb. 2008. Restructuring Research: Communication Costs and the Democratization of University Innovation. *American Economic Review* 98(4): 1578-1590.
- Agrawal, A., R. Henderson. 2002. Putting patents in context: Exploring knowledge transfer from MIT. *Management Science* 48(1): 44-60.
- Agrawal, A., I. Cockburn and J. McHale (2006), Gone But Not Forgotten: Labor Flows, Knowledge Spillovers, and Enduring Social Capital, *Journal of Economic Geography* 6, pp. 571-591.
- Alavi, M., D. Leidner. 2001. Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly* 25(1): 107-136.
- Alcacer, J., M. Gittelman. 2006. Patent citations as a measure of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics* 88(4): 774-779.
- Alcacer, J., M. Gittelman, and B. N. Sampat. 2009. Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy* 38(2): 415-427.
- Almeida, P., B. Kogut. 1999. Localization of Knowledge and the Mobility of Engineers in Regional Networks. *Management Science* 45(7): 905-917.
- Angrist, J. D., G. W. Imbens, and D. B. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91: 444-455.
- Atanasov, V. and B. Black. 2014. Shock-Based Causal Inference in Corporate Finance Research. Working Paper.
- Azoulay, P., J. Graff Zivin and B. Sampat (2011), The diffusion of scientific knowledge across time and space: Evidence from professional transitions for the superstars of medicine, NBER Working Paper 16683.
- Benner, M., J. Waldfoegel (2008), Close to you? Bias and precision in patent-based measures of technological proximity, *Research Policy* 37(9), pp. 1556-1567.
- Bloom, N., R. Sadun, and J. Van Reenen (2012), Americans do IT Better: US Multinationals and the Productivity Miracle, *American Economic Review* 102(1), pp. 167-201.
- Bloom, N., M. Schankerman and J. Van Reenen (2013), Identifying technology spillovers and product market rivalry, *Econometrica* 81(4), pp. 1347-1393.
- Bresnahan, T. and S. Greenstein. 1996. Technical progress and co-invention in computing and in the uses of computers. *Brookings Papers on Economic Activity, Microeconomics* 1996: 1-83.
- Brown, J. S. and P. Duguid. 1991. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science* 2(1): 40-57.
- Brynjolfsson, E. and L. Hitt (2003), Computing Productivity: Firm-Level Evidence, *The Review of Economics and Statistics*, 85(4), 793-808.
- Cassiman, B. and R. Veugelers (2002), R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium, *American Economic Review*, pp. 1169-1184.
- Cohen, W. and D. Levinthal (1989), Innovation and Learning: The Two Faces of R&D, *The Economic Journal* 99(397), 569-596.
- Cohen, W. and D. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35: 128-152.
- Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper 7552.

- Criscuolo, P. and B. Verspagen (2008), Does it matter where patent citations come from? Inventor vs. examiner citations in European Patents, *Research Policy* 37(10), 1892-1908.
- Dasgupta, P. and P. David. 1994. Toward a new economics of science. *Research Policy* 23: 487-521.
- Ding, W., S. Levin, P. Stephan, and A. Winkler (2010), The Impact of Information Technology on Academic Scientists' Productivity and Collaboration Patterns, *Management Science* 56(9), 1439-1461.
- Duguet, E., M. MacGarvie. 2005. How well do patent citations measure flows of technology? Evidence from French innovation surveys. *Economics of Innovation and New Technology* 14(5): 375-393.
- Feldman, M. P. and D. B. Audretsch. 1999. Innovation in cities: Science-based diversity, specialization, and localized competition. *European Economic Review* 43(2): 409-429.
- Fleming, L. (2001), Recombinant Uncertainty in Technological Search, *Management Science* 47(1), 117-132.
- Forman, C. 2005. The corporate digital divide: Determinants of Internet adoption. *Management Science* 51(4): 641-654.
- Forman, C., A. Goldfarb, and S. Greenstein. 2002. Digital dispersion: An industrial and geographic census of commercial Internet use. NBER Working Paper 9287.
- Forman, C., A. Goldfarb, and S. Greenstein (2005), How Did Location Affect the Adoption of the Commercial Internet? Global Village vs. Urban Density, *Journal of Urban Economics* 58(3), 389-420.
- Forman, C., A. Goldfarb, and S. Greenstein. 2008. Understanding the inputs into innovation: Do cities substitute for internet firm resources? *Journal of Economics and Management Strategy* 17(2): 295-316.
- Forman, C., A. Goldfarb, and S. Greenstein. 2012. The Internet and local wages: A puzzle. *American Economic Review* 102(1): 556-575.
- Forman, C. and N. van Zeebroeck (2012), From wires to partners: How the Internet has fostered R&D collaborations within firms, *Management science* 58(8), pp. 1549-1568.
- Furman, J. L., M. K. Kyle, I. Cockburn, and R. M. Henderson. 2005. Public & Private Spillovers, Location, and the Productivity of Pharmaceutical Research. *Annales D'Économie et de Statistique* 79/80: 167-190.
- Gao, G. and L. Hitt. 2004. IT and Product Variety: Evidence from Panel Data. *Proceedings of the 25th International Conference on Information Systems (ICIS)*, Washington, DC, 2004.
- Gaulé, P. and N. Maystre. 2011. Getting Cited: Does open access help? *Research Policy* 40: 1332-1338.
- Gaspar, J. and E. L. Glaeser. 1998. Information Technology and the Future of Cities. *Journal of Urban Economics* 43(1): 136-156.
- Gold, A., A. Malhotra and A. Segars (2001), Knowledge management: An organizational capabilities perspective, *Journal of Management Information Systems* 18(1), 185-214.
- Grant, R. (1996), Toward a Knowledge-Based Theory of the firm, *Strategic management journal* 17(S2), 109-122.
- Greenstein, S. and M. Mazzeo. 2014. The role of differentiation strategy in local telecommunication entry and market evolution: 1999-2002. *Journal of Industrial Economics* 54(3): 323-350.
- Griffith, R., S. Lee and J. Van Reenen (2007), Is distance dying at last? Falling home bias in fixed effects models of patent citations, NBER Working Paper 13338.
- Griliches, Z. 1990. Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28(4): 1661-1707.
- Hall, B., A. Jaffe, and M. Trajtenberg. 2005. Market Value and Patent Citations. *The RAND Journal of Economics* 36(1): 16-38.
- Hansen, M.T. 1999. The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge Across Organizational Subunits. *Administrative Science Quarterly* 44(1): 82-111.

- Hausman, J. 2001. Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *The Journal of Economic Perspectives* 15(4): 57-67.
- Henderson, R. and I. Cockburn. 1994. Measuring Competence? Exploring Firm Effects in Pharmaceutical Research. *Strategic Management Journal* 15: 63-84.
- Jaffe, A. (1986), Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value, *American Economic Review* 76(5), pp. 984-1001.
- Jaffe, A., 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.
- Jaffe, M. Trajtenberg, and M. S. Fogarty. 2000. The Meaning of Patent Citations: Report on the NBER/Case Western Reserve Survey of Patentees. *American Economic Review* 90(2): 215-218.
- Kim, H. 2012. The Effect of Free Access on the Diffusion of Scholarly Ideas. Working Paper.
- Kleis, L., P. Chwelos, R. Ramirez, and I. Cockburn (2012), Information technology and intangible output: The impact of IT investment on innovation productivity, *Information Systems Research* 23(1), pp. 42-59.
- Kogut, B. and U. Zander (1992), Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology, *Organization Science* 3(3), pp. 383-397.
- Kogut, B. and U. Zander. 1993. Knowledge of the Firm and The Evolutionary Theory of the Multinational Corporation. *Journal of International Business Studies* 24(4): 625-645.
- Lampe, R. 2012. Strategic citation. *Review of Economics and Statistics* 94(1): 320-333.
- Lawrence, S. 2001. Free Online Availability Substantially Increases a Paper's Impact. *Nature* 411(6837): 521.
- Leiponen, A. and C. Helfat. 2011. Location, Decentralization, and Knowledge Sources for Innovation. *Organization Science* 22(3): 641-658.
- Lychagin, S., J. Pinkse, M. Slade and J. Van Reenen (2010), Spillovers in space: Does geography matter?, NBER Working Paper 16188.
- MacGarvie, M. (2006), Do firms learn from international trade?, *Review of Economics and Statistics* 88(1), pp. 46-60.
- McAfee, A. 2009. *Enterprise 2.0: How to Manage Social Technologies to Transform Your Organization*. 2009. *Enterprise 2.0: How to Manage Social Technologies to Transform Your Organization*. Harvard Business School Press.
- McCabe, M. and C. M. Snyder. 2013. The Rich Get Richer and the Poor Get Poorer: The Effect of Open Access on Cites to Science Journals Across the Quality Spectrum. Working Paper.
- McDermott, R. (2000), Why information technology inspired but cannot deliver knowledge management, *California Management Review*, 41(4), 103-117.
- McElheran, K. 2014. Delegation in Multi-Establishment Firms: Evidence from I.T. Purchasing. *Journal of Economics and Management Strategy* 23(2): 225-258.
- Miller, D. J., M. J. Fern, and L. B. Cardinal. 2007. The Use of Knowledge for Technological Innovation within Diversified Firms. *The Academy of Management Journal* 50(2): 308-326.
- Nelson, R. and S. Winter (1982), *An evolutionary theory of economic change*, Cambridge (US), Harvard University Press., 1982.
- Orlando, M. (2004), Measuring spillovers from industrial R&D: on the importance of geographic and technological proximity, *RAND Journal of Economics* 35(4), 777-786.
- Pavitt, K. 1991. What makes basic research economically useful? *Research Policy* 20(2): 109-119.
- Penner-Hahn, J., J. M. Shaver. 2005. Does International Research and Development Increase Patent Output? An Analysis of Japanese Pharmaceutical Firms. *Strategic Management Journal* 26: 121-140.
- Roach, M., W. M. Cohen. 2013. Lens of Prism? Patent Citations as a Measure of Knowledge Flows from Public Research. *Management Science* 59(2): 504-525.
- Rosenkopf, L. and A. Nerkar (2001), Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry, *Strategic Management Journal* 22(4), pp. 287-306.

- Sambamurthy, V., R. W. Zmud. 1999. Arrangements for information technology governance: A theory of multiple contingencies. *MIS Quarterly* 23(2): 261-290.
- Singh, J. (2005), Collaborative Networks as Determinants of Knowledge Diffusion Patterns, *Management Science*, 51(5), pp. 756-770.
- Singh, J. 2008. Distributed R&D, cross-regional knowledge integration, and quality of innovative output. *Research Policy* 37: 77-96.
- Singh, J. and A. Agrawal (2011), Recruiting for ideas: How firms exploit the prior inventions of new hires, *Management Science* 57(1), pp. 129-150.
- Singh, J. and M. Marx (2013), Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity, *Management Science* 59(9), pp. 2056-2078.
- Stock, J. H. and M. Yogo. 2005. Testing for weak instrument in linear IV regression. D. W. K. Andrews, ed. *Identification and Inference for Econometric Models*. Cambridge University Press, Cambridge, UK, 80-108.
- Stuart, T. (2000), Interorganizational alliances and the performance of firms: A study of growth and innovation rates in a high-technology industry, *Strategic management journal* 21(8), 791-811.
- Stuart, T. and J. Podolny (1996), Local search and the evolution of technological capabilities, *Strategic Management Journal* 17(S1), 21-38.
- Teece, D. J. 1977. Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-How. *The Economic Journal* 87(346): 242-261.
- Teece, D. (1988), Capturing value from technological innovation: Integration, strategic partnering, and licensing decisions, *Interfaces* 18(3), 46-61.
- Vandenbosch, D. and M. J. Ginzberg. 1996. Lotus Notes and Collaboration: Plus ca change. *Journal of Management Information Systems* 13(3): 65-82.
- Veugelers, R. and B. Cassiman (2004), Foreign subsidiaries as a channel of international technology diffusion: Some direct firm level evidence from Belgium, *European Economic Review* 48(2), pp. 455-476.
- Zhao, M. Conducting R&D in Countries with Weak Intellectual Property Rights Protection. *Management Science* 52(8): 1185-1199.

Appendix

*Table 0: Descriptive statistics*

<b>Variable</b>	<b>Mean</b>	<b>StDev</b>	<b>Min</b>	<b>Max</b>	<b>N</b>
Nb 5-year citations	0.31	3.57	0.00	372.00	37720
Nb 5-year citations from collaboration	0.24	2.98	0.00	304.00	37720
At least one 5-year citation	0.06	0.23	0.00	1.00	37720
At least one 5-year citation from collaboration	0.05	0.21	0.00	1.00	37720
Basic Internet in both locations	0.20	0.40	0.00	1.00	37720
Log of per-establishment R&D spending	3.02	1.47	-0.46	7.73	37720
Log of patent stock in prior 10 years	2.84	1.64	0.00	9.02	37720
Log of patent stock in current period	1.78	1.48	0.00	7.61	37720
Technological proximity	0.09	0.18	0.00	1.00	37720
Technological proximity N.A.	0.33	0.47	0.00	1.00	37720
Log of establishment employees	7.69	1.13	5.30	12.04	37720
Share of local employment in manufacturing	0.20	0.06	0.04	0.52	37720
Local average weekly wages	543.60	85.44	306.48	848.33	37720
Log of local employment	13.84	0.94	10.35	15.70	37720
Log of number of local patents	6.69	1.21	0.69	9.13	37720