Internet adoption and knowledge diffusion*

Chris Forman
Dyson School of Applied Economics and Management
Cornell University
137 Reservoir Avenue
Ithaca, NY 14853
USA
chris.forman@cornell.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

November 2017
Preliminary and Incomplete

Abstract

Can ICT reduce the localization of knowledge? We investigate this question by studying the impact of basic Internet access on cross-location knowledge flows within the same firm. We construct a large data set of Internet adoption and patent citations among dyadic pairs of firm-locations between 1992-1998. We find that when both locations in the pair adopt basic Internet there is an increase in the likelihood of a citation between the citing and (potential) cited location, and that this likelihood increases more when the pair is working in similar research areas and when the research areas in the citing location are less specialized. These results are robust to a range of robustness analyses, including an instrumental variable strategy.

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

* We thank seminar participants at Ecoles des Mines de Paris, HEC Paris, University of California, Irvine, the Academy of Management Annual Meeting, the Digital Innovation Workshop at Boston College, and the European Conference on Information Systems for helpful comments and suggestions. 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). However, even within firms, it is often difficult to effectively transfer and use knowledge produced elsewhere within an organization (Allen 1977; Teece 1977). As a result, recent research has found somewhat mixed results about how the presence of multiple R&D locations influences the output and value of innovation with a firm (Leiponen and Helfat 2011, Singh 2008).

In this paper we study how digital innovation influences knowledge flows within firms.[1] The increasing digitization of the inputs into innovation—e.g., the documents used and communications that arise during innovation—has the potential to increase the flows of knowledge across geographically disparate organizations by lowering the costs of identifying useful information as well as the costs of transmitting it (e.g., Alavi and Leidner 2001, Kleis et al. 2012). However, the digitization of innovation may not be equally effective at facilitating all types of knowledge transfer. There are significant ex ante differences in the costs of transferring R&D-related knowledge, depending upon the nature of the knowledge, relationship between the sender and recipient, among other things.[2] Information technology systems may be better at reducing some of these costs than others, leading to a prediction of unequal efficacy of such systems in increasing knowledge transfer depending upon the context.

---

[1] Here we define digital innovation as in Nambisan et al. (2017): "the creation of (and consequent change in) market offerings, business processes, or models that results from the use of digital technology."

[2] The literature detailing the challenges to knowledge transfer is too extensive to review here. For two early examples, see Teece (1977) and Allen (1977).

At present, due to data constraints, we have little direct systematic empirical evidence of the effects of digitization of innovation inputs on knowledge flows within firms. This is particularly true for R&D knowledge flows, which may be difficult to absorb because of their specialized nature as they may require the combination of general and abstract knowledge with concrete information to produce novel inventions (Arora and Gambardella 1994). This is a significant gap in understanding. If digitization improves knowledge flows within firms, it has the potential to significantly reshape the traditional trade-offs associated with geographically dispersed R&D operations. Further, if there is unevenness in the circumstances over which ICT is able to reduce knowledge transfer costs, this has significant implications for the direction of inventive activity.

We examine the implications of the rapid declines in communication costs that occurred around the time of the initial commercialization of the Internet. Our research approach combines detailed data related to Internet adoption within firm establishments with that on patent citation patterns within large organizations. Our primary experiment examines changes in patent citations that occur between dyadic pairs of firm locations after both locations adopt Internet technology in the late 1990s. We label this type of digital innovation *dyadic Internet adoption.* We find that when dyadic Internet adoption occurs, it increases the likelihood of a patent citation between the firm locations by 1.2 percentage point, or 17.6%. This number excludes citations made to and by patents resulting from collaborations between the two locations.

Our results may be influenced by the presence of unobservables that could influence both the likelihood of adopting Internet technology and citation patterns within the firm. To address this potential concern, we instrument for Internet adoption using variables that will influence the costs of adopting Internet technology but which should have little impact on knowledge flows within firms. To further justify a causal interpretation, we provide additional evidence investigating where and when the effects of adoption of Internet technology arise.

We next investigate what types of knowledge flows are particularly influenced by the onset of lower communication costs. Adoption of Internet technology may lead users to tap in to more novel sources of knowledge that can lead to higher-impact inventions (Uzzi, Mukherjee, Stringer, and Jones 2013). However, knowledge that is dissimilar to a research organization's prior experience may also be harder to absorb and incorporate into new inventions (Cohen and Levinthal 1989, 1990). We find adoption of Internet technology leads to knowledge flows that reinforce existing areas of expertise. Specifically, 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. Further, we show that dyadic Internet adoption is less likely to lead to new knowledge flows when the receiving research group is narrowly specialized in their fields of research. Overall, these results show that the impact of Internet adoption on knowledge flows is uneven, and depends upon the ex-ante distribution of prior knowledge within the firm.

We investigate the underlying mechanism that gives rise to our results. Our results may reflect the ability of Internet technology to facilitate dyadic communication between sources and recipients of knowledge flows, as when a scientist provides advice and input when asked by a colleague. Alternatively, our results may reflect the creation of new –often browser-based– tools for searching digitized documents. To provide insights into which of these mechanisms shape our results, we examine differences in outcomes when both locations in the pair adopt to cases where only the citing location does so. Our results provide evidence that both mechanisms may be at work, and that knowledge flows under either mechanism are facilitated when there already exists a common knowledge base between firm locations.

1.1 Related Literature

Our findings contribute to several fields of research. First, we relate to a broad literature that has demonstrated that knowledge flows are localized (e.g., Jaffe et al. 1993; Thompson and Fox-Kean 2005).[3] In particular, we are most closely related to studies that have shown the benefits of co-location to knowledge exchange and innovation within organizations (Allen 1977), and that has demonstrated the impact of co-location on innovation, commercialization, and product quality outcomes (Catalini Forthcoming; Bercovitz and Feldman 2011; Gray, Siemsen, and Vasudeva 2015). We also relate to a line of work that investigates the efficacy of organizational practices that shape knowledge flows and knowledge exchange within geographically dispersed research organizations (e.g., Hansen 1999; Henderson and Cockburn 1994; Zhao 2006). However, this work has not illustrated the impact of IT investments on knowledge flows.

Our research is also related to a line of work that has investigated whether IT investments have changed the geography of scientific production.[4] [5]Early network technologies such as BITNET have been found to increase scientific collaboration, specifically across multiple universities (Agrawal and Goldfarb 2008; Ding et al. 2010; Walsh et al. 2000). However, research that has investigated the effects of IT investment on the benefits of co-location within

---

[3] For a recent review of this literature, see Feldman and Kogler (2010).

[4] Another line of works has investigated whether ICT investment is associated with an increase in innovation and productivity more broadly (e.g., Hall, Lotti, and Mairesse 2013; Kleis et al. 2012; Ravichandran et al. Forthcoming; Wu, Hitt, and Lu 2016).

[5] We further acknowledge a significant literature in the field of information systems on knowledge management (for one review see Alavi and Leidner 2001). While this research informs our thinking of the challenges of knowledge absorption, it does not centrally focus on the issues of digital innovation that are our central concern.

firms has thus far delivered somewhat mixed results. In a paper that is perhaps closest to this one, Forman and van Zeebroeck (2012) use a similar cross-industry data set to provide evidence that adoption of an early generation of Internet technology leads to an increase in collaborations between researchers located in distant geographic units within firm. However, recent work has shown that Internet investments may not reduce the benefits of co-location between manufacturing and R&D (Gray, Siemsen, and Vasudeva 2015).

As noted earlier, an additional goal of our research is to examine differential effects of digitization on knowledge flows based on the stock of knowledge at the citing location and overlap of knowledge between the potential source and recipient. In that way we speak to recent work that has argued that while new communications technologies increase collaboration and knowledge flows across physical space, they may also reduce interaction between individuals with differing tastes, preferences, and social/professional circles (Rosenblat and Mobius 2004, Van Alstyne and Bryjolfsson 2005). Some have even argued that this may result in the balkanization of science (Van Alstyne and Brynjolfsson 1996). While we are unable to confirm the specific data-generating processes argued for in these earlier papers, we do provide empirical evidence that supports the view that lower communication costs will differentially increase knowledge flows from researchers of similar fields relative to those who work in different research areas.

2 Research Framework

Our research objective is to identify how knowledge flows within organizations will change when the organization adopts a new technology that lowers communication costs. Knowledge flows are difficult to measure. We take the approach pioneered by Jaffe, Trajtenberg, and Henderson (1993) that measures knowledge flows through patent citations. Specifically, we examine self-citations of patents across locations within the same firm. The advantage of this approach is that organizations are legally obligated to disclose the knowledge that they build upon when they file a patent, and so citations offer a systematic way of tracing knowledge flows within organizations. Further, self-citations have been shown to be positively correlated with firm value (Hall, Jaffe, and Trajtenberg 2005) and so examining the factors that influence self-citations is economically important.

Adoption of new IT can lower communication costs through different mechanisms. One mechanism is by facilitating direct communication between users within an organization, as when an inventor asks a colleague for help (e.g., Allen 1977). Another mechanism arises when IT adoption lowers the costs of searching for now-digitized documents, as when new web browser tools are used to search for patents within the organization. Our data do not allow us to directly observe the process of information discovery within an organization, and so cannot directly identify between these two mechanisms. However, our research design will enable us to separately identify dyadic Internet adoption from that where only the citing location adopts. We will use this distinction to provide some insights into the relative importance of these different processes.

2.1 How does IT adoption improve dyadic communication?

As noted above, inventors can obtain knowledge through a number of sources. One is by accessing formal or informal documents within an organization, such as patents, publications, or internal technical reports. Another is by asking other employees directly. In this subsection we describe the second of these two process for knowledge acquisition and how it is influenced by Internet technology adoption. We discuss the process of search for internal documents in the next subsection.

The process for knowledge acquisition from other employees can be motivated by models of search (Garicano 2000). In our setting, a piece of knowledge found elsewhere in the organization may help the employee to solve a problem that arises within the research process. Employees continue to search so long as the benefits of search exceed the costs. As they continue to search, they may be able to identify a better solution to their problem. However, search is costly, and the outcome of future search is also uncertain. For one, additional search may not yield a new idea that is useful. Further, even if employees hear a solution to their problem they may not be able to apply it effectively if they do not have the requisite knowledge (Cohen and Levinthal 1989, 1990; Nelson and Winter 1982, Teece 1988).

Adoption of new information technology can reduce the search and transfer costs of accessing knowledge (Bolton and Dewatripont 1994, Garicano 2000). Our focus is on the adoption of Internet technology around the time of the initial commercialization of the Internet, and includes basic services such as the World Wide Web (including intranets) and SMTP-based email. If the combination of search and transfer costs fall sufficiently, then this increases the likelihood of a transfer of knowledge. This represents a joint hypothesis that adoption of Internet technology by both the sender and receiver will increase the likelihood of search as well as the likelihood of knowledge flow conditional on search.

The specific set of technologies that we study do not include more sophisticated systems to support collaborative work. In short, the type of technology that we study lowers the costs of bilateral communication, but is less well-suited to facilitating real-time interactions among a large group of dispersed workers. We consider this to be an advantage of our research design, as it limits the set of ways in which technology adoption can influence knowledge flows (and so reduces the set of potential underlying mechanisms).

2.2 IT adoption and technical document search

Another way of acquiring knowledge in an organization is by reading technical documents (Allen 1977). These can include formal documents that are published externally, such as patents or journal publications, or internal documents such as lab notes. Like the acquisition of knowledge from people, the acquisition of knowledge from documents can be viewed though a model of search. Users will continue searching for documents until the benefits no longer exceed costs. However, both the costs and expected benefits of an incremental search will be lower, relative to the case of searching from people.

As an example, consider changes to the search technology for patents. Prior to the widespread diffusion of the Internet, patent search tools were available on CD-ROMs. However, the CD-ROMs were infrequently updated, and costly to use. For example, there were 140 patents in the CD-ROM library for Japanese patents in 1994, so a patent search on any one CD would result in a search for less than 1% of patents (Martin 2016). Digitization of patent documents through a web interface allowed for faster searches, often with better tools that went beyond class-based search.

In short, adoption of Internet technology reduces the costs of accessing information that has already been codified. By providing access to real-time information on patents and other technical documents within the organization, it assists in locating knowledge that previously would be difficult to find. However, unlike the mechanism described in the prior section, use of Internet technology in this way will not facilitate digitization of ad hoc interactions that arise from questions about the underlying knowledge and how it may improve upon and interact with the components of an existing invention under development.

In the prior two subsections we described how adoption of Internet technology can reduce the costs of transferring knowledge within an organization. The costs of transferring knowledge across locations in a geographically dispersed organization are high (e.g., Teece 1977). As a result, we specifically investigate the impact of Internet adoption on knowledge flows between geographically dispersed establishments within the same firm.

We seek to provide insights into whether increases in knowledge flows arise because of declines in the costs of dyadic communication or because of the lower search costs of accessing existing documents. To do this, we compare how the effects on Internet adoption on citation behavior change when only the citing location adopts—which will reduce the search costs of accessing documents but not affect the costs of dyadic transfer—to the effects when both the citing and cited location adopt—which will facilitate the costs of dyadic communication.

We discuss issues around model identification more comprehensively in a later section, but discuss two particular issues here with the interpretation of our results. One concern is that adoption of Internet technology will increase all citation behavior, not just that between geographically dispersed establishments. To investigate that possibility, we compare the effects of Internet adoption on citation between geographically dispersed establishments from that between inventors who are in the same location.

Another issue arises because patents are both a measure of invention and a measure of appropriability (e.g., Arora, Ceccagnoli, and Cohen 2008). Thus, adoption of a new communication technology will both make it easier for inventors to locate and draw from knowledge from new locations (as described above), but can also make it easier for firms to identify related knowledge that makes it more effective in appropriating the value from its inventions. In our baseline analysis we will be unable to distinguish between these different mechanisms, but present analyses that suggest some of the effects of IT on knowledge flows will occur through the actions of inventors.

2.3 Conditions when the effects of ICT adoption on knowledge flows will be strongest

We next examine heterogeneity in our results. Our specific interest is related to how the respective knowledge bases of the source and recipient influence how dyadic Internet adoption will shape knowledge flows.

On the one hand, innovation is often spurred through the novel combination of ideas that give rise to new insights (e.g., Weitzman 1998, Jones 2009, Jones, Wuchty, and Uzzi 2008). Recent research has demonstrated that scientific research that draws upon multiple fields will have a greater impact (i.e., more likely to be cited in later research) (Uzzi, Mukherjee, Stringer, and Jones 2013). This suggests that knowledge flows that are derived from dissimilar fields have the potential to significantly increase the value of new inventions.

On the other hand, searching across different research domains is more costly and uncertain (Fleming 2001, Schilling and Green 2011). To start with, inventors in related fields may have stronger social ties that have been developed through prior collaborations. Stronger social ties will make it easier for inventors to ask questions and will also make it more likely they receive answers to questions. Further, when inventors recombine ideas from similar fields, they are able to draw upon past experience to determine the boundaries of successful regions of recombinant space (Fleming 2001).

Further, even conditional on identifying useful knowledge, inventors may experience challenges absorbing and putting to use new ideas from different areas (Cohen and Levinthal

1989, 1990; Henderson and Clark 1990). 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).

Based on these earlier findings, both the returns and costs of knowledge diffusion may be higher when the knowledge base of the source and recipient in the transfer are more dissimilar. This in turn will depend in part upon the breadth (or focus) of their respective knowledge bases (i.e. how specialized they are), and how they overlap (i.e. their proximity). Specifically, if research groups are specialized in one narrow area, they may benefit particularly from the opportunity to acquire knowledge from elsewhere. However, if specialization is higher, the knowledge base is narrower and hence it may be more difficult to absorb new knowledge (Cohen and Levinthal 1989, 1990). For similar reasons, if the source and recipients share a larger body of knowledge (i.e. their knowledge bases are more overlapping), transfer and absorption may be less costly, but the potential returns to knowledge diffusion may be lower as well.

Our main hypothesis is that the dyadic adoption of Internet technology will decrease the costs of searching for and transferring new knowledge. On the one hand, other things equal the (gross) benefits of Internet-enabled knowledge transfer will be greatest among dissimilar or more specialized knowledge bases for which the benefits of use are greatest. On the other hand, the type of Internet technology we consider will not make it easier for inventors to recombine or absorb new knowledge, even if helps them to access it. Nor will it help to establish network ties

(McAfee 2009). Internet adoption should therefore have little impact on the ability of inventors to absorb and put to use knowledge that is more distant from their own.

Because specialization and proximity in research will influence costs and benefits in opposite directions, we are unable to sign ex ante the net effects of these differences, and instead examine the data to determine them. We do therefore not assert ex ante hypotheses but rather allow the data to show the net effects of the mechanisms we describe here above. These net effects are important, as they speak to how IT shapes the direction of new invention within an organization.

3 Empirical strategy

We argue that adoption of basic Internet will be associated with a decline in the costs of knowledge transfer between two research locations within a firm. As a result, we expect an increase in knowledge flows between two firm locations who both adopt basic Internet. We follow a long line of prior literature starting with Jaffe, Trajtenberg, and Henderson (1993) and use citations between patents to measure knowledge flows.

In our research strategy we estimate the change in the incidence of citations between any two geographically-distant establishments within a given firm when the two firm locations adopt Internet technology, compared to two locations that do not adopt Internet technology. In the next section we describe our approach for measuring the average treatment effect of Internet adoption on citation flows between firm locations. We also discuss identification of the econometric model. We then demonstrate our approach for measuring how the effects of Internet adoption on citations is influenced by the specialization of research knowledge within the citing location and also by the proximity in research fields between the two locations

3.1 Estimating the average effect of Internet on citations

In our baseline estimation approach we used fixed effects panel data regression models. The unit of observation is a within-firm location dyad-year, and due to data constraints we observe behavior every other year. 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 10 preceding years (we have explored robustness to alternative time windows). X_{ijkt} is a vector of time-varying controls at the establishment-pair level such as the log of patent stock in the pair over the prior 10-year period and the log of per-establishment R&D spending. Z_{ijkt} is a vector of time-varying controls for local characteristics such as local average weekly wages and the log of local employment. Our main variable of interest is $Internet_{ijkt}$, which indicates whether locations *j* and *k* of firm *i* had both adopted Internet in year *t*. μ_{ijk} measures pair fixed effects and τ_t measures time fixed effects.

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 in Internet adoption across locations within firms; in general, firms did not adopt basic Internet across all establishments at the same time.

3.2 Model identification

A relationship between adoption of Internet technology and citations between locations does not, in itself, allow us to assert that our findings are consistent with the data-generating processes described above. Our research design faces challenges related to the multiple data-generating processes that could give rise to our results; the presence of unobserved variables that could be influencing our estimates; and measurement error in the key dependent and independent variables. We discuss these in the next subsections.

3.2.1 Differences in interpretation of results

The results from estimating the regression model in equation (1) could arise from multiple data-generating processes. One potential concern with the above specification is that patent citations between two firm locations may increase after Internet adoption because of an increase in collaborations between the two locations: Specifically, a positive coefficient on β may reflect increases in collaborations between locations j and k that are themselves precipitated by the adoption of Internet. 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%. If citations are more likely among collaborators, then our results may reflect an increase in collaborations between inventors in the pair. To address this concern, we identify patents that were co-invented between the two locations and exclude them from our measures of citations in our baseline specification. However, because this may induce concerns about bias arising from selection in our estimation sample, we have rerun all of our estimates including citations arising from collaborative patents and our results are qualitatively similar.

As noted earlier, increases in citations arising from Internet adoption can arise both from a decline in search and transfer costs from geographically dispersed dyadic communication, as well as a decline in the costs of searching for geographically dispersed documents through an improved search technology. To begin to identify between these alternative explanations, we re-estimate regression (1) adding an alternative version of our Internet variable which is equal to one whenever the citing location adopts Internet technology. When only the citing location adopts, this will decrease the costs of searching for geographically dispersed documents but will not influence the costs of electronic dyadic communication. Thus, if the effects of Internet adoption at the citing location influences citation behavior we will take this as evidence for the importance of Internet as a tool for accessing geographically dispersed knowledge assets.

Another possibility is that Internet adoption reduces the costs of accessing knowledge from everywhere in the organization, not just locations that are geographically dispersed. In this case, the localization of knowledge flows would not change. To investigate this possibility, we examine the effects of Internet adoption on the likelihood of a citation between inventors within the same MSA and compare it to our baseline estimates.

More broadly, a general concern with our research design is that there might be unobserved time-varying factors that could be correlated both with Internet adoption and with knowledge flows. For example, a firm-wide effort to increase knowledge flows between researchers could give rise to Internet adoption.

To address these concerns, we explore the robustness of our results to a range of a robustness tests. First, we explore whether our results appear at the “correct” time. We examine whether a variable capturing dyadic Internet adoption prior to when adoption actually occurs is associated with an increase in citations. If we observe that such a pre-adoption indicator is associated with citation flows, this is evidence that is suggestive of omitted variable bias.

Second, we explore the robustness of our results to the use of instrumental variables. We instrument for Internet adoption using variables that will shift the costs of adopting Internet technology but should have little effect on knowledge flows. Because the instruments are shaped by regional variation, we compute the variable for each of the two locations in the dyad and then take the average. We use two instruments. The first instrument captures local telecommunications costs: It is the year in which the state changed to rate of return (ROR) regulation for telecommunications services. Changes in regulatory policy 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.

3.2.2 Potential issues arising from measurement and model specification

Our model estimates may be subject to biases arising from measurement error and model specification. For one, we do not have a direct measure of knowledge flows, and use patent citations to capture knowledge flows. A commonly used assumption is that a citation from patent A to patent B reflects that patent A builds upon the knowledge in patent B (e.g., Jaffe, Trajtenberg, and Henderson 1993). We acknowledge the limitations of using patent citations as a proxy for knowledge flows (e.g., Alcacer and Gittelman 2006, Roach and Cohen 2013). In particular, not all citations reflect knowledge flows and some knowledge flows will not be reflected in citations. This problem is exacerbated in research designs that mix self-citations with citations to patents from other firms, which may be driven by different strategic imperatives. Our focus on organizational self-citations will mitigate some of these concerns. The key issue for identification is that firms are not simultaneously changing their self-citation behavior (e.g., as a research of a change in appropriability strategy) simultaneously with Internet investment. We discuss this possibility further in the context of our identification strategy and robustness analyses.

In our main analysis we focus upon a linear probability model rather than a binary choice model like a probit or logit for several reasons. First, we rely on within variation for identification and removing time invariant heterogeneity in a probit or logit model requires either stronger assumptions (in the case of the probit) or does not enable the computation of marginal effects (in the case of the logit). Further, the computation of marginal effects is less straightforward in nonlinear models, particularly for models including interaction terms (Ai and Norton 2003), and so our results using the linear probability model are easier to interpret. In general, we view our estimates as a linear approximation to an underlying nonlinear model.

Each firm location pair-year combination appears twice in our data, once with a given firm-location appearing as the citing location and once with it appearing as the cited location. To address standard error concerns related to duplicate values of the same covariates across multiple observations in our sample, we create an index for each firm location pair that is independent of the identity of the citing and cited location and cluster our standard errors around that index.

3.3 Exploring heterogeneity in the effects of Internet on citations

We next seek to understand how the effects of Internet adoption will vary based upon the similarity in research domains between the source and recipient of the knowledge flows. To do this, we interact our measure of Internet adoption with measures of technological proximity in research areas across locations in the dyad and a measure of the specialization of the research domain of the recipient.

For example, to examine whether the effects of Internet adoption on citation behavior is systematically different for locations working in similar research areas, we estimate the following regression equation:

$$Citation_{ijkt} = \alpha_1 X_{ijkt} + \alpha_2 Z_{ijkt} + \beta Internet_{ijkt} + \gamma Internet_{ijkt} \times \Pr oximity_{ijkt} + \mu_{ijk} + \tau_t + \varepsilon_{ijkt}$$

(2)

Where *Proximity(ijkt)* measures the technological proximity of locations j and k of firm i at time t. Our models for estimating differences in the effects of Internet based on the citing location's specialization in research are estimated similarly.

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). Our estimation sample is from 1992-1998.

4.1 Patent Data

Within each firm, we then use citations between patents invented at different locations as a proxy for within-firm cross-location knowledge spillovers. To do so, we use data on patents filed by multi-establishment US manufacturing firms at the United States Patent and Trademark Office (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 a patent with application date t from location j to another patent invented in location k over the previous ten-year period. Our focus on the extensive margin of whether there exist any citations—rather than the count of the number of citations between locations—in part reflects the distribution of our dependent variable: only 8.4% of dyads have a citation between them (6.7% when collaborative patents are excluded). This low number is partly explained by a share (29%) of locations having no patents in a given year. Conditional on whether the source location has at least one patent in the focal year, 12% of dyads have at least one citation between them (9.5% with collaborative patents excluded). However, we experimented with using the number of citations as our dependent variable within a series of Poisson count data regressions, and our results are robust.

Our analysis requires us to identify both the firm and location in which a patent is invented.[6] 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 within-firm location dyad-years, with separate observations based on the citing location in the 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.

4.2 Information Technology Data

Our data on Internet adoption comes from a private 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

[6] The discussion in this section and the next is a summary of the relevant issues. For further details, see Forman and van Zeebroeck (2012).

firm. The data have been used in a wide range of studies related to the adoption of IT (e.g., Bresnahan and Greenstein 1996, Forman, Goldfarb, and Greenstein 2005) and IT productivity (e.g., 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; Forman, Goldfarb, and Greenstein 2015).

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); or the establishment reports using the Internet for research purposes. We assume that no establishments has 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 70% had adopted by 1998.

The CI data are collected at the establishment level. We match establishments to MSAs using the establishment zip code to match our establishment-level IT data to USPTO patent data. Whenever we have several establishments within a given MSA, we take the average of all variables, except for Internet adoption where we consider that a location has adopted as soon as one establishment within the location has. To obtain controls from COMPUSTAT such as R&D expenses, we further match the firm identifier in the CI database to a COMPUSTAT GVKEY.

4.3 Firm-MSA Pairs and dependent variable

The focus of our study is to examine the effects of IT investments on cross-establishment knowledge flows within organizations. We estimate 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 *j* (invented during the focal period *t*) that cites a patent invented at location *k* over the ten-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 (*j*-*k* and *k*-*j*). The dependent variable is different in the two configurations: it will reflect citations by patents invented in location *j* to patents invented in location *k* in the first case, and citations from *k*-patents to *j*-patents in the second. Our main dependent variable is a binary measure indicating whether establishment *j* makes at least one citation to a patent invented at establishment *k* in the preceding 10-year period. Because we want to ensure that citations are not just reflecting collaborative projects between *j* and *k*, we build an alternative measure of our dependent variable that excludes citing and cited patents that were co-invented at the two locations (*j* and *k*). Our alternative dependent variable is therefore a binary measure indicating whether location *j* makes at least one reference to a patent invented at location *k*, outside of co-invented patents. This alternative dependent variable will represent our baseline measure.

We restrict our estimation sample to firm MSA dyad-year combinations 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 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. Because we do not observe employment in 1992 and 1994 (we do not have CI data for these years), we assume 1996 values for these years. 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. Specifically, the proximity between locations j and k of firm i is computed as:

$$Prox_{ijk} = \frac{\sum_{c=1}^{C} P_{ijct} P_{ikct}}{\sqrt{\left(\sum_{c=1}^{C} P^2{}_{ijct}\right)}\left(\sum_{c=1}^{C} P^2{}_{ikct}\right)}$$

Where C is the total number of technological classes considered, and t is the period over which we compute proximity. Following Benner and Waldfogel (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. We computed proximity over the two-year period (1989-1990) that precedes our analysis period (1991-1998). For some firm-location pairs this variable was undefined because one of the establishments in the pair had no patents during the period considered. In this case we added a dummy variable to indicate that proximity is undefined. In our regressions, rather than the nominal proximity score, we use a dummy that equals to 1 if the proximity score of the focal dyad is above the median in our sample and zero otherwise.

In a similar way, we also compute the degree of technological specialization of each establishment using a Herfindahl index. The resulting score ranges between 1 (all patents are concentrated in a single class) and 0 (all patents belong to different classes). Here again, we compute the specialization index based on the distribution of the focal establishment's patents across 525 US classes prior to the start of our sample period.[7] If s_{ijt} is the share of patents applied for by establishment *j* in period *t* that fall in the i^{th} technological class, the Herfindahl index is then given by:

[7] Our main specification uses 3-digit US classes (525 technological classes), which we believe is a well-balanced level of measurement for specialisation at the firm level. We report, however, consistent results using Hall, Jaffe and Trajtenberg's classification (referred to as "HJT"), which includes 37 sub classes. See van Zeebroeck et al. (2006) for a discussion of the different parameters in specialisation measures based on patent data.

$$H_{jt} = \sum_{i=1}^{n} s_{ijt}^{2}$$

Both measures, technological proximity among pairs and technological specialization at the establishment level, can be measured at different points in time: simultaneously with the focal observation period (running in-sample measure), over the entire period of analysis (fixed in-sample measure), or over a period preceding our analysis period (pre-sample measure). We believe that the pre-sample measure is the most reliable as it reduces potential endogeneity concerns (e.g. if technological specialization and proximity are themselves influenced by Internet adoption). We however report results using different versions of these measures. They are robust.

We further control for other potential sources of heterogeneity. We control for gross innovative output of the dyad and the environment using patent stocks at the dyad and county levels. To account for potential changes in patenting productivity, we compute the dyad's patent stock in the current period (running count) in addition to the stock observed over a moving 10-year period (i.e., cumulative stock of patents invented at any of the two locations over the previous 10 years). We include these measures to ensure that our results are not explained by raw increases in patent output. Additional county-level controls include the share of local employment in manufacturing, local average weekly wages, and the log of local employment.

Table 1 reports descriptive statistics for our main variables.

5 Empirical results

We first establish the relationship between Internet adoption and citation flows. We explore the robustness of these results to a variety of robustness tests, including a falsification exercise and an instrumental variables approach. We next explore how the effect of Internet on knowledge flows differ based on the overlap in research domains between the source and recipient. Last, by comparing differences in results when one location versus both locations in a pair adopt basic Internet, we provide insights preliminary insights in to the mechanism underlying our results.

5.1 Baseline results

We begin by establishing a baseline result between Internet adoption and dyadic citation patterns by estimating regression equation (1). The first two columns show results using a version of the dependent variable that includes citations that arise from patents that involve a collaboration between inventors in the two locations. We focus on data points arising from firm-location pairs for which we have data for all four data point in our sample, but include results from an unbalanced panel as a robustness check. The coefficient associated with Internet adoption is positive and statistically significant (at the 10% level), suggesting that Internet does foster knowledge flows between distant establishments within firms: if both locations adopt Internet, this translates into a 1 percentage point increase in the likelihood of a citation between them. Given that the mean likelihood of a within-pair citation is 8% in our sample, this implies that Internet adoption increases the incidence of citations by almost 12.5% in relative terms. In column 2, we explore the robustness of our results to the use of an unbalanced panel. The point estimate is lower in this baseline specification and no longer significant at conventional thresholds. As often happens in linear probability models (e.g., Athey and Stern 2002; Agrawal and Goldfarb 2008), the within R^2 value is low.

As noted above, the results in the first two columns of Table 2 will reflect both the effects of the Internet on promoting knowledge flows through collaborations as well as the effects of the Internet on promoting knowledge flows conditional on collaboration. In columns 3 and 4 we rerun regression model (1) using an alternative version of our dependent variable that excludes citations that arise from patents in which there is a collaboration between inventors across the two locations in the firm MSA pair. The point estimate is qualitatively consistent with our baseline (reported in the first column), but larger, and now significant at the 5% level with both balanced and unbalanced panels.

The point estimate implies that when both establishments in the pair adopt basic Internet the likelihood of a citation between the pair increases 1.2 percentage points, or 17.9%. Going forward, to more carefully isolate the effects of Internet on knowledge flows, we will focus on our results excluding collaborative patents. However, we have re-estimated all of our models using citations that include collaborations and the results are qualitatively similar.

Table A1 explores the robustness of our main results. Columns 2 and 3 of Table A1 investigates the timing of our results. We first include in column 3 a measure that is 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. As it appears in column 2, our lead measure of Internet adoption does not have any effect on citations. 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 in our data given that our baseline measure includes citations over a two-year period (based upon application date). It could alternatively reflect a lack of power in our statistical test (many firm locations adopt Internet at the end of our sample period). Columns 4 and 5 of Table A1 show that our results are robust to the use of two different 5-year windows: citations to more recent patents (invented at the cited location within the last five years) or to older patents (invented at the cited location between six and ten years ago). Adoption of Internet technology increases citations to both older and more recent patents.

One question is whether adoption of Internet technology lead to an increased number of citations in general, not just those across establishments within the same firm. To explore this possibility, in Appendix Table A2 we present results of a regression that explores the effects of Internet adoption on patent citations within the same location at the firm. As in our baseline analysis, we do this using two different approaches—using a measure of citations that includes those arising from collaborations between inventors in the same location and another measure that excludes citations between inventors in the same location (i.e., single-inventor citations). As a benchmark, the likelihood of observing a citation to a same-MSA patent including those arising from collaborations is 46.5%; the likelihood of observing a citation to a same-MSA patent excluding those arising from collaborations is 29.6%. In all of our specifications, adoption of Internet technology is not associated with an increase in the likelihood of observing a citation from patents invented in the same MSA.

5.2 Instrumental Variable Estimates

To further address concerns about omitted variable bias, we present the results of instrumental variables estimates. We explore the implications of the wo instruments that were described above (year of first change to ROR regulation and number of ARPANET nodes), each of which will influence the costs of adopting basic Internet. Table 3 presents the second stage results; we present first stage results in Appendix Table A3.

We include the results for both instruments separately and then together. 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 effects of Internet adoption. 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 our baseline estimates (reproduced in column 1 of Table 3 for convenience) is 0.0118, while the same coefficient ranges in size from 0.133 (for first change to ROR regulation) to 0.2848 (for number of ARPANET nodes). One potential reason for this pattern is that 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.

5.3 Exploring the Effects of Technological Proximity and Specialization

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 and when the citing location is technologically specialized. Our identification assumptions are somewhat weaker in this section than earlier. Here our primary assumption is that there do not exist unobserved variables that are correlated with Internet and citations and that are differentially trending for pairs where there is high proximity and where knowledge is specialized. Table 4 shows the results of estimating model 2 and its analog for the specialization tests.

Column 2 shows heterogeneity in the effects of Internet based on the proximity of research areas between the establishments in the pair. The results show that the effects of Internet are much stronger for establishments who are in similar fields. Our proximity measure requires that both locations in the pair patent at least once in the pre-sample (1989-1990) period over which proximity is calculated; For some pairs of locations there were no patents over this period with which to compute the proximity measure. We were unable to compute a proximity measure for 45.1% of observations in our sample. To control for the presence of location pairs for which we are unable to compute a proximity measure, we include an addition interaction between Internet and an indicator that proximity is not available. The point estimate for our variable capturing Internet x Proximity pre-sample is above median is 0.0463, implying that establishment pairs who adopt Internet and who are technologically proximate see a 5.13 percentage point (=0.00500 + 0.0463) increase in the likelihood of a citation. In contrast, pairs who adopt Internet who are not technologically proximate see no increase in the likelihood of observing a citation.

We have probed the robustness of this result in several ways. First, we have computed our proximity measure using in-sample data. We have both allowed proximity to vary in-sample and also created a time-invariant version of the variable computed over our entire period of analysis (1991-1998). Our results are robust to both approaches. Also, one alternative explanation for our results is that technological proximity between two locations is proxying for prior collaborations between the locations. That is, locations that are technologically proximate will have also collaborated before, and such locations will benefit disproportionately from Internet adoption because of prior ties between the locations will increase the likelihood of a citation, other things equal (Singh 2005). To address this possibility, in Appendix Table A4 we present results from regressions that include proximity, presence of prior collaborations, and both in the same regression. We find that when the interactions of Internet and both proximity and prior collaborations are included in the same regression, the statistical and economic significance of the Internet and proximity interaction remains similar to that as in Table 4 while there is no effect of Internet on citation flows for pairs of locations that have collaborated before.

Column 3 provides the results of our proximity interactions using instrumental variables. To construct the instruments, we use the same two instruments as in column 4 of Table 3 and interact them with our proximity measures (proximity is above the median and proximity value is above the median). Thus, we have six instruments in total: the two instruments in table 3 along with their interactions with proximity and proximity not available. 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 the OLS results in column 2).

In column 4 we show the results of regressions that allow the marginal effect of Internet to vary based on the specialization of research interests in the establishment. As described above, our baseline measure of specialization is based upon 525 US classes. As was the case for our proximity calculations, there are some citing locations for which there were no patents in the pre-sample period with which to calculate a measure of specialization. As a result, we include an additional term – and its interaction with Internet—which is equal to one when we are unable to compute a specialization measure.[8] This happens for 25% of our sample.

---

[8] As noted earlier, our sample only include firm-locations that patent during our sample period. However, this is a within-sample condition. As a result, a firm-location could be included in our sample and still not patent during the pre-sample period and therefore lack a measure of specialization.

Column 4 shows that the effects of Internet adoption decrease significantly when the citing location is very specialized in its research. When the citing establishment is not specialized, pairs of locations who adopt Internet see a statistically significant (at the 1% level) 3.2 percentage point increase in the likelihood of a citation. When the citing establishment is specialized, there is no statistically significant effect of Internet on the likelihood of observing a citation. We have explored the robustness of this result to different ways of measuring specialization. Specifically, we have computed specialization based on in-sample data using the US class data. We have also computed specialization using the more aggregated Hall-Jaffe-Trajtenberg (Hall, Jaffe, and Trajtenberg 2001) technological areas. Our results are robust to all of these changes. Column 5 also shows that our results are robust to the use of instrumental variables.[9]

Finally, columns 6 and 7 report additional estimates that include proximity and specialization in the same regression. The results are qualitatively similar to when their effects are estimated separately. In additional robustness analyses that are not shown, we also explored a three-way interaction between Internet, proximity, and specialization but found that proximity and specialization have independent effects on behavior.

5.4 Exploring the Effects of dyadic and browser-based search

As noted earlier, our results so far could reflect dyadic communication between inventors doing work related to the citing and cited patents, improved search capabilities that occur after Internet adoption that do not require dyadic communication (e.g., browser-based search), or a

[9] As was the case for proximity, we use our two instruments in Table 3 and interact those with our measures of specialization and specialization not available.

combination of the two processes. Unfortunately, we do not directly observe the behavior of inventors in our data and so we cannot conclusively identify the underlying mechanism that gives rise to our results. However, we can provide suggestive evidence about the relative importance of these two mechanisms by observing the impact of Internet adoption when one and both locations in the pair adopt. If both locations in the pair adopt, then our results may be due either to improve dyadic communication or improved search capabilities. However, if only one location adopts, then it must be because of improved document search capabilities only.

Table 5 presents results where only the citing location and both locations adopt. In column 1 of Table 5 we provide results when only the citing location adopts. The coefficient estimate on citing location adoption (0.0083) is positive and statistically significant at the 5 percent level, however it is lower than our baseline estimate presented in column 3 of Table 2 when both locations adopt (0.0118). Of course, when citing Internet is included alone it may be capturing variance related to dyadic adoption, and so in column 2 we present estimates when both the citing and cited locations adopt. In this case, we can see that the average effect of citing Internet adoption across our entire sample is not statistically or economically significant; however there is a statistically and economically significant incremental effect when the cited location adopts in addition to the citing location (0.0102; significant at the 10% level). The combined effects when both the citing and cited locations adopt (i.e., dyadic adoption) is 0.0128 and statistically significant at the 5% level.

In column 3 we explore heterogeneity in these effects when both the citing and cited locations adopt. The individual point estimates of the incremental effects of Internet adoption at the citing (0.0049) or both (0.0011) locations in the absence of technological proximity in the citing and cited locations are not statistically significant at conventional levels. Similarly, the incremental effects of Internet adoption when proximity is present are also not statistically significant. However, to identify the dyadic effects of Internet adoption in the presence of proximity we must take linear combinations, and these show that adoption in the presence of proximity can lead to statistically and economically significant changes in behavior. When the citing location adopts Internet technology and there is also technological proximity, the likelihood of a citation between locations increases by 2.33 percentage points (=0.0011+0.0222; significant at the 10% level). When the cited location adopts in addition to the citing location, the likelihood of a citation increases by an additional 3.11 percentage points (=0.0262+0.0049; significant at the 10% level). Overall, when both locations adopt and there is technological proximity between establishments, the likelihood of a citation increases by 5.44 percentage points (significant at the 1% level). In short, these results show that adoption of Internet technology at the citing and cited locations will have an effect on citation behavior only when the two locations are technologically proximate.

Column 4 shows how the results of Internet adoption depend on the extent to which the citing location is technologically specialized. When the citing location is not specialized, the effects of Internet adoption will lead to increases in citation flows through both dyadic- and search-based channels. When only the citing location adopts and the citing location is not specialized, Internet adoption will lead to a 1.55 percentage point increase in the likelihood of a citation (significant at the 10% level). When the cited location also adopts and the citing location is not specialized, this leads to an additional 1.83 percentage point increase in the likelihood of a citation after Internet adoption (significant at the 10% level). In sum, when both the citing and cited locations adopt Internet and the citing location is specialized, this leads to a 3.38 percentage point increase in the likelihood of a citation (significant at the 1% level). In contrast, when the citing location is specialized, neither adoption at the citing or cited location (or both) influences citation behavior.

Column 5 provides results when the effects of both proximity and specialization are included in the same regression. The results show that when specialization is low and proximity is high, both adoption at the citing establishment (marginal effect 0.0275; significant at the 10% level) and citing+cited establishment (marginal effect 0.0600; significant at the 1% level) have a significant effect on citation behavior. However, when specialization is high, adoption of Internet at the citing establishment does not influence behavior, even in the presence of proximity. However, adoption of Internet at both the citing and cited establishments does lead to an increase in citations when proximity is high, even when specialization is present.

We attempted to rerun the models above using our instrumental variables approach. However, given the large number of potentially endogenous variables in these models, we were unable to capture the independent effects of citing and cited internet adoption.

6 Discussion and implications

We analyze the impact of Internet adoption on the diffusion of technological knowledge within multi-establishment manufacturing firms in the US. Our results show that adoption of an early form of Internet technology significantly increases the likelihood of observing a patent citation between geographically dispersed R&D locations within the same firm. Our results are robust to the use of a timing falsification, and we find no evidence that Internet adoption increased citation flows within the same location of the firm. Further, our results are robust to the use of an estimation strategy that uses variables that shift the cost of adoption as instruments for our (potentially endogenous) Internet adoption variable.

We find that the effects of Internet adoption on knowledge flows are stronger in circumstances where the citing establishment is less specialized in its knowledge base and when the proximity in research fields between the citing and cited establishment is greater. Last, we begin to probe whether the benefits from Internet adoption arise from lower costs of dyadic communication between scientists or improved document search capabilities. Our results show that there are significant incremental benefits that arise from dyadic communication, relative to the case when costs of searching for documents decline due to the adoption of Internet at the citing establishment. However, proximity and specialization influence the benefits of adoption at citing and cited establishments in similar ways.

Our results show that Internet adoption will increase knowledge flows between locations only when they share a common knowledge base. These results have implications for our understanding of the impact of digitization on the rate and direction of inventive activity its impact on economic growth. The recombination of novel ideas is a key element of new growth theory. As noted by Paul Romer, "Economic growth occurs whenever people take resources and rearrange them in ways that make them more valuable" (Romer 2008). It is widely believed that IT investment and the digitization of economic activity can play an important role in this recombination (Romer 2008, Brynjolfsson and McAfee 2014). However, our results suggest that at least early forms of Internet technology reinforced existing connections among vertical groups of technologies, rather than facilitating recombination of ideas across technology areas.

These results may reflect the limitations of the type of Internet technology that we study in overcoming the costs of searching across technological areas. For example, users who conduct document searchers using browser-based technology may search using technology classes or even keywords that are based on their pre-existing knowledge of the patent classification system.

Further, Internet-enabled applications may be effective at facilitating distant communication but less effective at establishing social connections (Gaspar and Glaeser 1998) that are important for search based on dyadic communication between a sender and receiver of knowledge. These limitations may reflect in part the characteristics of the specific sets of Internet-enabled technologies that we study in this paper. For example, recent patent search tools allow search approaches that are more flexible than that based on keywords and patent classes (Martin 2016; Huc 2017). Further, more recent technologies such as online forums may be more effective at facilitating one-to-many communication patterns and allow knowledge searchers to access information from users with whom they may not have existing social connections (McAfee 2009). On the other hand, the same filtering technologies that allow more flexible searches may encourage researchers to more narrowly define search fields and so further reinforce reliance on fields where the researcher has expertise.

Another issue is that while Internet adoption may have decreased the costs of transferring knowledge, it may have had little impact on the ability of scientists to absorb knowledge from dissimilar fields and deploy it in new applications.

This perspective has been emphasized in prior research related to knowledge management (e.g., Alavi and Leidner 2001) and absorptive capacity (Cohen and Levinthal 1989, 1990) which have emphasized the need for prior related knowledge to absorb and deploy new knowledge received in new circumstances in useful ways. The inability of organizations to deploy knowledge in new ways 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 &

Levinthal, 1990). Kogut and Zander (1992) have suggested that absorbing knowledge from a different field is difficult due to differences in shared codes between the different groups involved, while Leonard-Barton (1988) emphasized the necessity of a mutual adaptation between the source and recipient of knowledge transfers.

Increasingly entities like InnoCentive have enabled broadcast search that allows the firm to draw on the expertise of the crowd both within and between firms (e.g., Jeppesen and Lakhani 2010). While such firms may be effective at overcoming previous technological challenges in accessing knowledge from diverse fields, technology alone will not resolve the challenges of knowledge absorption that have been highlighted in the literature. Intermediaries like InnoCentive play a role in framing problems for their clients, such work may play a role in overcoming the potential barriers identified in this paper.

In short, this paper has identified a fact: that IT-enabled knowledge flows are greater between inventors who are working in related fields. It has identified a range of potential reasons for this fact, ranging from factors that inhibit knowledge search and transfer across more distant fields to those that may inhibit absorption. It remains to be seen whether later generations of IT will inhibit or reinforce these barriers. We leave this as a question for future research.

Our analysis is subject to limitations. For one, as noted above, we study a specific set of Internet technologies over a specific time period. The impact of the adoption of Internet technologies on knowledge flows may be different over a later time period. Second, while the nature of our data are unique in that they enable us to study the implications of a change in communication costs on knowledge flows over a large and diverse sample, our unit of analysis makes it difficult for us to pin down exactly the mechanism through which Internet adoption changes behavior. Last and related, we have pursued a variety of analyses to provide a causal interpretation of our results. However, we are also limited by the nature of our data in the types of analyses that we can perform. We look forward to future research that investigations these questions in other settings using other data.

Another limitation is that our research focuses primarily on intra-firm knowledge flows. Future work should examine how the adoption of IT influences knowledge flows between institutions such as firms, universities, or research institutes. Research on the latter entities would provide new insights into the ‘balkanization of science’ hypothesis. Our focus on firms means that scientists and inventors must focus their research efforts on areas that contribute to the broader product strategy of the firm. Thus, while our results can certainly be influenced by barriers to search, transfer, and absorption costs across distant fields, they are less likely to be influenced by purposeful changes in research interests and connections enabled by technology. However, researchers in universities, for example, will have academic freedom on where to focus their research efforts, and so the risks that IT investments will facilitate changes in search strategies based on preferences will be significant (Rosenlat and Mobius 2004). We hope our findings encourage future research in this important area.

References

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, J. McHale. 2006. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. Journal of Economic Geography 6: 571-591.

Agrawal, A., A. Goldfarb. 2008. Restructuring Research: Communication Costs and the Democratization of University Innovation. *American Economic Review* 98(4): 1578-1590.

Ai, C., E. C. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80(1): 123-129.

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.

Allen, T. J. 1977. *Managing the Flow of Technology*. Cambridge: MIT Press.

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, D. B. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91: 444-455.

Arora, A., A. Gambardella. 1994. The changing technology of technological change: general and abstract knowledge and the division of innovative labour. *Research Policy* 23: 523-532.

Arora, A., M. Ceccagnoli, W.M. Cohen. 2008. R&D and the patent premium. *International Journal of Industrial Organization* 26(5): 1153-1179.

Athey, S., S. Stern. 2002. The impact of information technology on emergency health care outcomes. *RAND Journal of Economics* 33(3): 399-432.

Benner, M., J. Waldfogel. 2008. Close to you? Bias and precision in patent-based measures of technological proximity, *Research Policy* 37(9): 1556-1567.

Bercovitz, J., M. Feldman. 2011. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy* 40: 81-93.

Bloom, N., R. Sadun, J. Van Reenen. 2012. Americans do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review* 102(1): 167-201.

Bolton, P., M. Dewatripont. 1994. The Firm as a Communication Network. *Quarterly Journal of Economics* 109(4): 809-839.

Bresnahan, T., 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.

Brynjolfsson, E., L. Hitt. 2003. Computing Productivity: Firm-Level Evidence. *The Review of Economics and Statistics* 85(4): 793-808.

Brynjolfsson, E., A. McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Machines*. Cambridge: MIT Press.

Catalini, C. Forthcoming. Microgeography and the Direction of Inventive Activity. *Management Science.*

Cohen, W., D. Levinthal. 1989. Innovation and Learning: The Two Faces of R&D. *The Economic Journal* 99(397): 569-596.

Cohen, W., D. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35: 128-152.

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.

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.

Feldman, M. P. and D. F. Kogler. 2010. Stylized Facts in the Geography of Innovation. In *Handbook of the Economics of Innovation, Volume 1*, eds. B. H. Hall and N. Rosenberg. Amsterdam: Elsevier, pp. 381-410.

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, 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, 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, S. Greenstein. 2015. Information Technology and the Distribution of Inventive Activity. In *The Changing Frontier: Rethinking Science and Innovation Policy*, eds. Adam Jaffe and Ben Jones, University of Chicago Press.

Forman, C., N. van Zeebroeck. 2012. From wires to partners: How the Internet has fostered R&D collaborations within firms, *Management Science* 58(8): 1549-1568.

Garicano, L. 2000. Hierarchies and the Organization of Knowledge in Production. *Journal of Political Economy* 108(5): 874-904.

Gaspar, J., E. L. Glaeser. 1998. Information Technology and the Future of Cities. *Journal of Urban Economics* 43(1): 136-156.

Greenstein, S., M. Mazzeo. 2006. The role of differentiation strategy in local telecommunication entry and market evolution: 1999-2002. *Journal of Industrial Economics* 54(3): 323-350.

Grey, J., E. Siemsen, G. Vasudeva. 2015. Colocation Still Matters: Conformance Quality and the Interdependence of R&D and Manufacturing in the Pharmaceutical Industry. *Management Science* 61(11): 2760-2781.

Griliches, Z. 1990. Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28(4): 1661-1707.

Hall, B., A. Jaffe, M. Trajtenberg. 2001. The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools. NBER Working Paper 8498.

Hall, B., A. Jaffe, M. Trajtenberg. 2005. Market Value and Patent Citations. *The RAND Journal of Economics* 36(1): 16-38.

Hall, B., F. Lotti, J. Mairesse. 2013. Evidence on the impact of R&D and ICT investments on innovation and productivity in Italian firms. *Economics of Innovation and New Technology* 22(3).

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.

Henderson, R. and K. B. Clark. 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly* 35(1): 9-30.

Henderson, R. and I. Cockburn. 1994. Measuring Competence? Exploring Firm Effects in Pharmaceutical Research. *Strategic Management Journal* 15: 63-84.

Huc, Olivier. 2017. The Challenges of Patent Knowledge Management – part 1. Blog Post, Available at https://www.linkedin.com/pulse/challenges-patent-knowledge-management-part-1-olivier-huc/.

Jaffe, A., M. Trajtenberg, R. Henderson. 1993 Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics* 108(3): 577-598.

Jeppesen, L.B., K.R. Lakhani. 2010. Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science* 21(5): 1016-1033.

Jones, B. F. 2009. The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *The Review of Economic Studies* 76(1): 283-317.

Kleis, L., P. Chwelos, R. Ramirez, I. Cockburn. 2012. Information technology and intangible output: The impact of IT investment on innovation productivity. *Information Systems Research* 23(1): 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.

Leonard-Barton, D. 1988. Implementation as mutual adaptation of technology and organization. Research policy 17(5): 251-267.

Leiponen, A. and C. Helfat. 2011. Location, Decentralization, and Knowledge Sources for Innovation. *Organization Science* 22(3): 641-658.

MacGarvie, M. 2006. Do firms learn from international trade? *Review of Economics and Statistics* 88(1): 46-60.

Martin, J. 2016. The Third Age of Patent Search Tools. Blog Post, Available at https://www.innography.com/blog/post/the-third-age-of-patent-search-tools.

McAfee, A. 2009. *Enterprise 2.0: How to Manage Social Technologies to Transform Your Organization*. Boston: Harvard Business School Press.

Miller, D. J., M. J. Fern, L. B. Cardinal. 2007. The Use of Knowledge for Technological Innovation within Diversified Firms. *The Academy of Management Journal* 50(2): 308-326.

Nambisan, S., K. Lyytinen, A. Majchrzak, M. Song. 2017. Digital innovation management: Reinventing innovation management research in a digital world. *MIS Quarterly* 41(1): 223-238.

Nelson, R., S. Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

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.

Ravichandran, T., S. Han, S. Mithas. Forthcoming. Mitigating Diminishing Returns to R&D: The Role of Information Technology in Innovation. *Information Systems Research.*

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.

Romer, P. 2008. Economic Growth. Library of Economics and Liberty. Available at http://www.econlib.org/library/Enc/EconomicGrowth.html.

Rosenblat, T.S. and M.M. Mobius. 2004. Getting Closer or Drifting Apart? *Quarterly Journal of Economics* 119(3): 971-1009.

Schilling, M.A. and E. Green. 2011. Recombinant search and the breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy* 40: 1321-1331.

Singh, J. 2005. Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science* 51(5): 756-770.

Singh, J. 2008. Distributed R&D, cross-regional knowledge integration, and quality of innovative output. *Research Policy* 37: 77-96.

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.

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.

Thompson, P., M. Fox-Kean. 2005. Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *American Economic Review* 95(1): 450-460.

Uzzi, B., S. Mukherjee, M. Stringer, B. Jones. 2013. Atypical Combinations and Scientific Impact. *Science* 342(6157): 468-472.

Van Alstyne, M., E. Brynjolfsson. 1996. Could the Internet Balkanize Science? *Science* 274(5292): 1479-1480.

Van Alstyne, M., E. Brynjolfsson. 2005. Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. *Management Science* 51(6): 851-868.

Van Zeebroeck, N., B. van Pottelsberghe, W. Han. 2006. Issues in measuring the degree of technological specialisation with patent data. *Scientometrics* 66(3): 481-492.

Walsh, J. P., S. Kucker, N. G. Maloney, S. Gabbay. 2000. Connecting minds: Computer-mediated communication and scientific work. *Journal of the Association for Information Science and Technology* 51(14): 1295-1305.

Weitzman, M.L. 1998. Recombinant Growth. *Quarterly Journal of Economics* 113(2): 331-360.

Wu, L., L. Hitt, and B. Lu. 2016. Data Analytics Skills, Innovation, and Firm Productivity. The Wharton School Research Paper No. 86. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2744789.

Zhao, M. 2006. Conducting R&D in Countries with Weak Intellectual Property Rights Protection. Management Science 52(8): 1185-1199.

Table 1 - Descriptive statistics

Variable	N	Mean	SD	Min	Max
Pair has at least one 10-year citation flow	37,720	0.08	0.28	0.00	1.00
Pair has at least one 10-year citation flow collaborative patents excluded	37,720	0.07	0.25	0.00	1.00
Basic Internet adopted in both locations	37,720	0.20	0.40	0.00	1.00
Log of per-establishment R&D spending	37,720	3.02	1.47	-0.46	7.73
Log of patent stock over previous 10 years	37,720	2.84	1.64	0.00	9.02
Log of patent stock in current period	37,720	1.78	1.48	0.00	7.61
Log of establishment employees	37,720	7.69	1.13	5.30	12.04
Share of local employment in manufacturing	37,720	0.20	0.06	0.04	0.52
Local average weekly wages	37,720	543.60	85.44	306.48	848.33
Log of local employment	37,720	13.84	0.94	10.35	15.70
Log of number of local patents	37,720	6.69	1.21	0.69	9.13
Year of change to ROR regulation x after 1996 dummy	37,720	994.78	994.79	0.00	1999.00
Mean number of ARPANET nodes x after 1996 dummy	37,720	0.63	1.43	0.00	9.00
Year of change to Price cap regulation x after 1996 dummy	37,720	47.19	47.21	0.00	99.00
Continuous proximity in research among two locations	37,720	0.09	0.18	0.00	1.00
Continuous proximity variable not applicable	37,720	0.33	0.47	0.00	1.00
Fixed pre-sample proximity between the 2 establishments (in 1990)	37,720	0.06	0.18	0.00	1.00
Average pre-sample specialisation (in 1990) based on USCL	37,720	0.34	0.35	0.00	1.00
Average pre-sample specialisation (in 1990) based on HJT	37,720	0.39	0.35	0.00	1.00
Average pre-sample specialisation (in 1990) not applicable	37,720	0.26	0.44	0.00	1.00

Table 2, Baseline

	Including collaborations	Including collaborations - Unbalanced panel	Baseline (excluding collaborations)	Baseline - Unbalanced panel
	(1)	(2)	(3)	(4)
Basic Internet in both locations	0.0100*	0.0066	0.0118**	0.0104**
	(0.0052)	(0.0051)	(0.0047)	(0.0046)
Log of per-establishment R&D spending	0.0021	0.0035	0.0048	0.0050
	(0.0055)	(0.0047)	(0.0050)	(0.0042)
Log of patent stock over previous 10 years	0.0358***	0.0339***	0.0251***	0.0231***
	(0.0038)	(0.0033)	(0.0034)	(0.0029)
Log of patent stock in current period	0.0282***	0.0271***	0.0222***	0.0220***
	(0.0021)	(0.0018)	(0.0019)	(0.0016)
Log of establishment employees	-0.0304***	-0.0257**	-0.0269***	-0.0243**
	(0.0109)	(0.0105)	(0.0099)	(0.0095)
Share of local employment in manufacturing	0.3329	0.2792	0.2731	0.1557
	(0.2419)	(0.2098)	(0.2220)	(0.1895)
Local average weekly wages	0.0007***	0.0007***	0.0005***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log of local employment	-0.1853***	-0.1816***	-0.1458***	-0.1339***
	(0.0541)	(0.0474)	(0.0513)	(0.0449)
Log of number of local patents	-0.0210	-0.0018	0.0018	0.0117
	(0.0140)	(0.0125)	(0.0127)	(0.0114)
R^2	0.03	0.03	0.03	0.03
N	37,720	51,340	37,720	51,340

Notes: The dependent variable is the incidence of a patent citation between a citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 3, Instrumental Variables, Second Stage

	Baseline	First change to ROR	Nb ARAPNET nodes	Both instruments
	(1)	(2)	(3)	(4)
Basic Internet in both locations	0.0118**	0.1333*	0.2848	0.1495*
	(0.0047)	(0.0804)	(0.2680)	(0.0783)
Log of per-establishment R&D spending	0.0048	0.0027	-0.0001	0.0024
	(0.0050)	(0.0054)	(0.0072)	(0.0054)
Log of patent stock over previous 10 years	0.0251***	0.0253***	0.0256***	0.0253***
	(0.0034)	(0.0034)	(0.0038)	(0.0035)
Log of patent stock in current period	0.0222***	0.0211***	0.0197***	0.0209***
	(0.0019)	(0.0020)	(0.0031)	(0.0020)
Log of establishment employees	-0.0269***	-0.0311***	-0.0364**	-0.0317***
	(0.0099)	(0.0107)	(0.0149)	(0.0108)
Share of local employment in manufacturing	0.2731	0.3265	0.3931	0.3336
	(0.2220)	(0.2338)	(0.2800)	(0.2351)
Local average weekly wages	0.0005***	0.0005***	0.0005***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log of local employment	-0.1458***	-0.1335**	-0.1183*	-0.1319**
	(0.0513)	(0.0541)	(0.0644)	(0.0544)
Log of number of local patents	0.0018	-0.0042	-0.0118	-0.0050
	(0.0127)	(0.0138)	(0.0189)	(0.0137)
Overidentification test (p-value)	N/A	N/A	N/A	0.357
Hausman test (p-value)				
N	37,720	37,720	37,720	37,720

Notes: The dependent variable is the incidence of a patent citation between a citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 4, Proximity and Specialization

	Baseline	With proximity pre-sample	IV - Both instruments - Proximity pre-sample interaction	With specialization pre-sample	IV - Both instruments - Specialization pre-sample interaction	With proximity & specialization pre-sample	IV - Both instruments - Proximity & specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Basic Internet in both locations	0.0118**	0.0050	0.1110	0.0318***	0.1802**	0.0152**	0.1365*
	(0.0047)	(0.0054)	(0.0766)	(0.0068)	(0.0773)	(0.0070)	(0.0785)
Internet X Proximity pre-sample is above median		0.0463***	0.1092***			0.0413***	0.0940***
		(0.0110)	(0.0197)			(0.0111)	(0.0200)
Internet x Proximity not available in 1990 data		-0.0083	-0.0116			0.0014	-0.0030
		(0.0058)	(0.0124)			(0.0074)	(0.0142)
Internet X Specialization above median				-0.0262***	-0.0767***	-0.0174**	-0.0564***
				(0.0070)	(0.0127)	(0.0070)	(0.0130)
Internet X Specialization (USCL) not applicable				-0.0432***	-0.0912***	-0.0284***	-0.0533***
				(0.0075)	(0.0146)	(0.0089)	(0.0154)
Log of per-establishment R&D spending	0.0048	0.0044	0.0016	0.0049	0.0028	0.0045	0.0019
	(0.0050)	(0.0050)	(0.0054)	(0.0050)	(0.0053)	(0.0050)	(0.0054)
Log of patent stock over previous 10 years	0.0251***	0.0280***	0.0314***	0.0260***	0.0272***	0.0274***	0.0308***
	(0.0034)	(0.0035)	(0.0037)	(0.0034)	(0.0034)	(0.0034)	(0.0037)
Log of patent stock in current period	0.0222***	0.0229***	0.0225***	0.0237***	0.0243***	0.0236***	0.0241***
	(0.0019)	(0.0019)	(0.0020)	(0.0019)	(0.0021)	(0.0019)	(0.0021)
Log of establishment employees	-0.0269***	-0.0232**	-0.0226**	-0.0254***	-0.0272***	-0.0229**	-0.0211*
	(0.0099)	(0.0099)	(0.0109)	(0.0098)	(0.0105)	(0.0098)	(0.0108)
Share of local employment in manufacturing	0.2731	0.2627	0.3065	0.2404	0.2465	0.2485	0.2595
	(0.2220)	(0.2213)	(0.2336)	(0.2207)	(0.2293)	(0.2207)	(0.2302)
Local average weekly wages	0.0005***	0.0004***	0.0003***	0.0004***	0.0004***	0.0004***	0.0003**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log of local employment	-0.1458***	-0.1375***	-0.1169**	-0.1361***	-0.1158**	-0.1348***	-0.1129**
	(0.0513)	(0.0507)	(0.0524)	(0.0509)	(0.0528)	(0.0506)	(0.0520)
Log of number of local patents	0.0018	0.0027	-0.0018	0.0024	-0.0022	0.0031	-0.0002
	(0.0127)	(0.0127)	(0.0134)	(0.0127)	(0.0135)	(0.0127)	(0.0133)
R^2	0.03	0.03		0.03		0.03	
N	37,720	37,720	37,720	37,720	37,720	37,720	37,720

Notes: The dependent variable is the incidence of a patent citation between a citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 5, Separating the effects of dyadic and browser-based search

	Only citing-location adoption	Both adoption measures	Proximity interactions	Specialization interactions	Full model
	(1)	(2)	(3)	(4)	(5)
Basic Internet in both locations		0.0102*	0.0049	0.0183*	0.0073
		(0.0059)	(0.0073)	(0.0104)	(0.0105)
Internet adoption at the citing location in the pair	0.0083**	0.0026	0.0011	0.0155*	0.0095
	(0.0041)	(0.0052)	(0.0063)	(0.0090)	(0.0090)
Internet X Proximity pre-sample is above median			0.0262		0.0252
			(0.0174)		(0.0177)
Internet x Proximity not available in 1990 data			-0.0011		0.0062
			(0.0095)		(0.0118)
Internet at citing location X Proximity fixed is above median			0.0222		0.0180
			(0.0152)		(0.0153)
Internet at citing location X Proximity not available			-0.0080		-0.0055
			(0.0084)		(0.0105)
Internet X Specialization above median				-0.0091	-0.0041
				(0.0121)	(0.0121)
Internet X Specialization (USCL) not applicable				-0.0200	-0.0157
				(0.0135)	(0.0156)
Internet at citing location X Specialization is above median				-0.0187*	-0.0145
				(0.0109)	(0.0108)
Internet at citing location X Specialization not applicable				-0.0252**	-0.0139
				(0.0118)	(0.0135)
Log of patent stock over previous 10 years	0.0250***	0.0251***	0.0289***	0.0262***	0.0281***
	(0.0034)	(0.0034)	(0.0035)	(0.0034)	(0.0035)
Log of per-establishment R&D spending	0.0048	0.0048	0.0041	0.0049	0.0042
	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
Log of patent stock in current period	0.0222***	0.0222***	0.0231***	0.0241***	0.0240***
	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Log of establishment employees	-0.0268***	-0.0269***	-0.0230**	-0.0251**	-0.0225**
	(0.0099)	(0.0099)	(0.0099)	(0.0098)	(0.0098)
Share of local employment in manufacturing	0.2684	0.2725	0.2548	0.2286	0.2358
	(0.2221)	(0.2219)	(0.2212)	(0.2200)	(0.2202)
Local average weekly wages	0.0005***	0.0005***	0.0004***	0.0004***	0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log of local employment	-0.1455***	-0.1455***	-0.1331***	-0.1319***	-0.1292**
	(0.0513)	(0.0513)	(0.0507)	(0.0508)	(0.0505)
Log of number of local patents	0.0016	0.0016	0.0025	0.0018	0.0027
	(0.0128)	(0.0128)	(0.0127)	(0.0127)	(0.0127)
R^2	0.03	0.03	0.03	0.03	0.03
N	37,720	37,720	37,720	37,720	37,720

Notes: The dependent variable is the incidence of a patent citation between a citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table A1, Robustness

	Baseline	Includes Internet Adoption Leads	Includes Internet Adoption Lags	5-Year Citations	6-10-Year Citations
	(1)	(2)	(3)	(4)	(5)
Basic Internet in both locations	0.0118**	0.0125**	0.0119**	0.0078*	0.0094**
	(0.0047)	(0.0060)	(0.0047)	(0.0046)	(0.0039)
Future Internet adoption		0.0007			
		(0.0038)			
Lagged Internet adoption			0.0042		
			(0.0088)		
Log of per-establishment R&D spending	0.0048	0.0048	0.0049	0.0031	0.0115***
	(0.0050)	(0.0050)	(0.0050)	(0.0047)	(0.0044)
Log of patent stock over previous 10 years	0.0251***	0.0251***	0.0251***	0.0314***	-0.0058***
	(0.0034)	(0.0034)	(0.0034)	(0.0032)	(0.0019)
Log of patent stock in current period	0.0222***	0.0222***	0.0222***	0.0193***	0.0083***
	(0.0019)	(0.0019)	(0.0019)	(0.0018)	(0.0011)
Log of establishment employees	-0.0269***	-0.0269***	-0.0266***	-0.0171*	-0.0256***
	(0.0099)	(0.0099)	(0.0099)	(0.0090)	(0.0088)
Share of local employment in manufacturing	0.2731	0.2731	0.2739	0.1914	0.3706**
	(0.2220)	(0.2220)	(0.2221)	(0.1950)	(0.1774)
Local average weekly wages	0.0005***	0.0005***	0.0005***	0.0003***	0.0007***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log of local employment	-0.1458***	-0.1457***	-0.1456***	-0.1100**	-0.1345***
	(0.0513)	(0.0513)	(0.0513)	(0.0469)	(0.0455)
Log of number of local patents	0.0018	0.0018	0.0019	0.0102	0.0128
	(0.0127)	(0.0128)	(0.0127)	(0.0119)	(0.0137)
R^2	0.03	0.03	0.03	0.02	0.04
N	37,720	37,720	37,720	37,720	37,720

Notes: The dependent variable is the incidence of a patent citation between a citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table A2, Same MSA

	Including collaborations – Different MSAs	Including collaborations - Same MSA	Including collaborations - Same MSA (Unbalanced)	Baseline	Baseline - Same MSA	Baseline - Same MSA (unbalanced)	IV Reg - Same MSA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Basic Internet in both locations	0.0114**	-0.0116	-0.0087	0.0130***	0.0232	0.0183	0.2044
	(0.0052)	(0.0200)	(0.0198)	(0.0048)	(0.0176)	(0.0175)	(0.2773)
Log of per-establishment R&D spending	0.0209***	0.0670***	0.0724***	0.0191***	0.0652***	0.0642***	0.0641***
	(0.0056)	(0.0207)	(0.0194)	(0.0051)	(0.0183)	(0.0177)	(0.0186)
Log of establishment employees	-0.0298***	0.0356	0.0352	-0.0265***	0.0513*	0.0521*	0.0393
	(0.0109)	(0.0276)	(0.0276)	(0.0099)	(0.0280)	(0.0281)	(0.0325)
Share of local employment in manufacturing	0.3546	0.2135	0.1821	0.2935	0.4339	0.3800	0.4570
	(0.2453)	(0.8209)	(0.8145)	(0.2243)	(0.6323)	(0.6299)	(0.6440)
Local average weekly wages	0.0008***	-0.0005	-0.0005	0.0005***	-0.0006	-0.0006	-0.0007
	(0.0001)	(0.0005)	(0.0005)	(0.0001)	(0.0004)	(0.0004)	(0.0004)
Log of local employment	-0.1415***	0.1362	0.1604	-0.1129**	-0.0439	-0.0638	0.0030
	(0.0545)	(0.1838)	(0.1810)	(0.0514)	(0.1623)	(0.1645)	(0.1724)
Log of number of local patents	-0.0110			0.0093			
	(0.0141)			(0.0128)			
R^2	0.02	0.03	0.03	0.02	0.02	0.02	-0.01
N	37,720	5,045	5,237	37,720	5,045	5,237	5,034

Notes: The dependent variable is the incidence of a patent citation between patents within the same MSA. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses. Log of number of local patents is not available in same MSA regressions. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table A3, Instrumental Variables, First Stage

	First change to ROR	Nb ARAPNET nodes	Both instruments
	(1)	(2)	(3)
Year of change to ROR regulation x after 1996 dummy	-0.0116***		-0.0115***
	(0.0017)		(0.0017)
Mean number of ARPANET nodes x after 1996 dummy		0.0064**	0.0059**
		(0.0029)	(0.0030)
Log of per-establishment R&D spending	0.0204**	0.0196*	0.0218**
	(0.0100)	(0.0101)	(0.0101)
Log of patent stock over previous 10 years	-0.0009	-0.0027	-0.0017
	(0.0058)	(0.0058)	(0.0058)
Log of patent stock in current period	0.0085***	0.0090***	0.0082***
	(0.0030)	(0.0030)	(0.0030)
Log of establishment employees	0.0376*	0.0355*	0.0381*
	(0.0205)	(0.0205)	(0.0205)
Share of local employment in manufacturing	-1.1310**	-0.5563	-1.2316**
	(0.4945)	(0.4995)	(0.4968)
Local average weekly wages	-0.0003	-0.0002	-0.0005*
	(0.0002)	(0.0002)	(0.0003)
Log of local employment	-0.0528	-0.0282	0.0133
	(0.1039)	(0.1076)	(0.1080)
Log of number of local patents	0.0470*	0.0370	0.0353
	(0.0262)	(0.0266)	(0.0266)
appyear==1994	-0.0100	-0.0068	-0.0081
	(0.0083)	(0.0084)	(0.0084)
appyear==1996	23.1322***	0.1033***	22.9093***
	(3.4534)	(0.0192)	(3.4558)
appyear==1998	23.7239***	0.6907***	23.5087***
	(3.4548)	(0.0337)	(3.4571)
R^2	0.61	0.61	0.61
F-statistic	44.48	4.69	24.26
Stock-Yogo Critical Values	8.96	8.96	11.59
N	37,720	37,720	37,720

Notes. First-stage dependent variable is an indicator for whether both MSAs in the pair have basic Internet. All regressions include time dummies. Stock and Yogo (2005) critical values are reported for maximal instrumental variable size >15%, respectively. Robust standard errors, clustered on firm-location pairs, are in parentheses.* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Appendix Table A4 - **Robustness check with citations including collaborations**

	Baseline - With proximity	Baseline - With prior collaborations, 1989-1990	Baseline - With prior collaborations, Pre-1991	Baseline - With proximity + prior collaborations, 1989-1990	Baseline - With proximity + prior collaborations, Pre-1991
	(1)	(2)	(3)	(4)	(5)
Basic Internet in both locations	0.0050	0.0087*	0.0091**	0.0046	0.0050
	(0.0054)	(0.0046)	(0.0046)	(0.0055)	(0.0054)
Internet X Proximity pre-sample is above median	0.0463***			0.0433***	0.0458***
	(0.0110)			(0.0110)	(0.0112)
Internet x Proximity not available in 1990 data	-0.0083			-0.0084	-0.0084
	(0.0058)			(0.0058)	(0.0058)
Internet X Establishments had collaborated in 89-90		0.0368*		0.0127	
		(0.0192)		(0.0197)	
Internet X Establishments had collaborated before 1991			0.0267		0.0018
			(0.0174)		(0.0181)
Log of patent stock over previous 10 years	0.0280***	0.0256***	0.0255***	0.0281***	0.0280***
	(0.0035)	(0.0034)	(0.0034)	(0.0035)	(0.0035)
Log of per-establishment R&D spending	0.0044	0.0049	0.0049	0.0044	0.0044
	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
Log of patent stock in current period	0.0229***	0.0223***	0.0223***	0.0229***	0.0229***
	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Log of establishment employees	-0.0232**	-0.0255**	-0.0256***	-0.0230**	-0.0232**
	(0.0099)	(0.0099)	(0.0099)	(0.0099)	(0.0099)
Share of local employment in manufacturing	0.2627	0.2577	0.2612	0.2576	0.2620
	(0.2213)	(0.2224)	(0.2221)	(0.2217)	(0.2215)
Local average weekly wages	0.0004***	0.0005***	0.0005***	0.0004***	0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log of local employment	-0.1375***	-0.1402***	-0.1403***	-0.1359***	-0.1372***
	(0.0507)	(0.0511)	(0.0511)	(0.0507)	(0.0506)
Log of number of local patents	0.0027	0.0020	0.0021	0.0028	0.0028
	(0.0127)	(0.0127)	(0.0127)	(0.0127)	(0.0127)
R^2	0.03	0.03	0.03	0.03	0.03
N	37,720	37,720	37,720	37,720	37,720

Notes: The dependent variable is the incidence of a patent citation between a citing firm-MSA and cited firm-MSA in the pair. All regressions include a constant term, firm-pair fixed effects, and time dummies. Robust standard errors, clustered on firm-location pairs, are in parentheses. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.