From wires to partners: How the Internet has fostered R&D collaborations within firms *

Chris Forman
College of Management
Georgia Institute of Technology
800 West Peachtree St. NW
Atlanta, GA 30308
USA
chris.forman@mgt.gatech.edu

Nicolas van Zeebroeck
ECARES
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

January 2011

Abstract

How did the diffusion of the Internet influence research collaborations within firms? We examine the relationship between business use of basic Internet technology and the size and geographic composition of industrial research teams between 1992 and 1998. We find robust empirical evidence that basic Internet adoption is associated with increased growth of citation-weighted collaborative patents in geographically dispersed teams. On the contrary, we find no evidence of such a link between Internet adoption and within-location collaborative patents, nor do we find any evidence of a relationship between basic Internet and single-inventor patents. We interpret these results as evidence that adoption of basic Internet significantly reduced the coordination costs of research teams, but find little evidence that a drop in the costs of shared resource access significantly improved research productivity. We further find that the benefits of Internet adoption are particularly strong for teams among research locations with larger economies of scale and scope, among those who had not collaborated before, among those who were active in highly collaborative fields, as well as among those who worked across dispersed research areas.

Keywords: R&D organization, geography of innovation, internet adoption, IT investments

* We thank Avi Goldfarb, Shane Greenstein, Marco Ceccagnoli, and participants at the ZEW-KMRC Workshop on Social Software, the EPFL-ETHZ Conference on Knowledge in Organizations, the 8th ZEW Conference on the Economics of IT, the 2010 International Conference in Information Systems (ICIS), the ICT and Economic Growth Conference (Munich) and seminar participants at ECARES, the University of Pennsylvania, the University of Washington, and the University of Minnesota for helpful comments and suggestions. We thank Michael Cha for invaluable research assistance. We gratefully acknowledge funding support from the General Motors Strategy Center at the Kellogg School of Management and the Software Industry Center at Carnegie Mellon University. Chris Forman acknowledges funding from a Sloan Industry Studies Fellowship from the Alfred P. Sloan Foundation. Nicolas van Zeebroeck acknowledges funding from FNRS. We also thank Harte Hanks Market Intelligence for supplying data. All opinions and errors are ours alone.

1. Introduction

An increasing fraction of scientific research is no longer done by individual inventors but by collaborative research teams (e.g., Adams et al. 2005; Hicks and Katz 1996; Jones 2009; Wuchty, Jones, and Uzzi 2007). This shift toward collaborative research is thought to be caused in part by increasing incentives for researchers to specialize, due to the increasing knowledge burden faced by scientists as scientific knowledge accumulates over time (Jones 2009). More broadly, increased specialization and division of labor among researchers may improve productive efficiency independent of the human capital investments of researchers.

Historically, collaborative work has been hampered by the existence of significant coordination costs that increase with team size, geographic dispersion, and heterogeneity of team composition (e.g., Becker and Murphy 1992). It is widely believed that by lowering these coordination costs, adoption of information technology (IT) such as the Internet may increase the returns to collaborative work (e.g., Cairncross 1997; Friedman 2005). However, while a small body of recent research has examined the implications of IT investment for collaborative academic research (Agrawal and Goldfarb 2008; Ding et al. 2010) to our knowledge there has been little systematic empirical work on the implications of IT investment for industrial research. This is a surprising gap in understanding. Collaborative research has not only been shown to be increasing in frequency but has also been shown to be more highly cited (e.g., Presser 1980; Sauer 1988). Further, since collaborative ties are known to increase the likelihood of knowledge flows (e.g., Singh 2005; Fleming et al. 2007), changes in collaboration patterns have important implications for the diffusion of knowledge within firms.

In this paper we take a first step toward empirically evaluating how IT investments shape research collaborations within firms. We motivate our hypotheses using prior models of team-based knowledge work, in particular the models of Becker and Murphy (1992) and Adams et al. (2005) that view optimal team size as a tradeoff between the benefits of specialization and division of labor versus increased coordination costs. We use this to motivate a set of hypotheses about how a decline in coordination costs will increase collaborative research output.

To test these hypotheses, we focus on the role of investments by firms in a set of Internet technologies that lower communications costs. We label the margin that we examine basic Internet; Prominent examples of basic Internet include Internet access or an internal intranet. The set of technologies we examine require little adaptation or co-invention (Bresnahan and Greenstein 1996) to be used successfully, and so allow us to focus on the short run changes to collaboration patterns that are made in response to a decline in communication costs.

We identify IT investments using a data set compiled by Harte Hanks Market Intelligence, a market research firm. As has been discussed elsewhere (e.g., Forman, Goldfarb, and Greenstein 2005), this data set represents among the best sources of information on the IT investments of private firms available.[1] To evaluate the implications of IT investment for research collaborations within firms, we combine this business IT investment data with data on US patenting activity from the US Patent and Trademark Office (USPTO). We link IT data collected at the establishment level to the number of patents invented within a metropolitan statistical area (MSA). Thus, in our analyses we examine variance over time in the number of patents invented by researchers in pairs of firm-MSA locations and among researchers within a given firm-MSA.

Our econometric approach examines the impact of Internet adoption on the number of patents invented by collaborating researchers within a firm (i.e., patents with more than one inventor). We compare the number of collaborative patents invented by researchers before basic Internet technology diffused (i.e., in 1992) to the number of patents after its diffusion (i.e., in 1998). That is, we use a difference-in-difference econometric estimation approach to identify the relationship between basic Internet investments and the pattern of research collaborations. Our sample period addresses a time period over which Internet technology had diffused but before enough time had elapsed for firms to change the geographic location of its researchers. This research design enables us to isolate the short run effects of a change in coordination costs, holding the internal organization and human capital decisions of the firm's researchers fixed.

Our first set of results assumes that basic Internet adoption is exogenous to research collaborations. We examine collaborations within pairs of heterogeneous geographically distant firm locations, where coordination costs are likely to be greatest. Our results show that when two locations within a firm both adopt basic Internet technology, the collaborative patents invented by researchers in both locations increase significantly compared to an otherwise identical pair without basic Internet. In contrast, we find that adoption of basic Internet has no impact on the number of collaborative patents among researchers within a single firm location, nor does it have any impact on the number of patents invented by lone inventors. We find that both results remain robust to numerous specifications and changes to controls. Together, we take this as evidence that by lowering coordination costs, basic Internet has increased the productivity of larger, geographically dispersed research teams relative to other types of research collaborations. While basic Internet technology may have increased researcher productivity in other ways—for example, by lowering access costs to shared resources—we find no evidence that these potential benefits resulted in increasing research output among research teams (including lone inventors) where ex ante coordination costs were low.

[1] A large number of researchers have now used these data to examine IT adoption (e.g., Bresnahan and Greenstein 1996) and productivity (e.g., Brynjolfsson and Hitt 2003; Bloom et al. 2009).

We next address the assumption that Internet adoption is exogenous. We first utilize the timing of Internet adoption as the source of a falsification exercise. We find no evidence that cross-location research collaborations prior to the diffusion of the commercial Internet (1990-1994) were correlated with the location-pair's later adoption of Internet technology in 1998.We also demonstrate that our results are robust to the use of instrumental variables. We employ two sets of instruments. The first instrument captures differences in the cost of Internet adoption across locations due to local regulatory conditions. The second uses industry-level variance in the benefits to Internet adoption.

We then motivate and test a series of comparative statics predictions for when the impact of basic Internet investments will be greatest. We find that basic Internet adoption has the greatest impact among researchers who had not collaborated before—and for whom ex ante coordination costs were likely high. Further, we find that Internet adoption had a significantly greater impact among those with researchers for whom the benefits of task specialization are likely to be large: namely, those who worked ex ante in highly collaborative research fields and in a highly dispersed set of research areas. In short, our results suggest that by lowering coordination costs, adoption of Internet technology has facilitated task specialization and division of labor in research. We also find that our results are strongest among firm-location pairs that were high patenters prior to the diffusion of the Internet (i.e., among those who were already research productive). We find that these results are also robust to the use of instrumental variables.

Our research contributes to a better understanding of the costs and benefits of scientific research collaborations, and in particular the implications of the diffusion of IT for collaborative work. Several of our findings differ significantly from that of prior work on the implications of IT investment for academic research collaborations.[2] In particular, one paper related to ours is Agrawal and Goldfarb (2008), who show that adoption of an earlier communication technology, BITNET, facilitated cross-institution collaboration in the academe, particularly among researchers in the same geographic region. In contrast, we examine a different setting: industrial research collaborations, and find that adoption of basic Internet was associated with a disproportionate increase in cross-location collaborations, with little effect on within-location collaborations. As we discuss in further detail below, we speculate that these results are due to differences in the way that firm and academic research collaborations are formed, the nature of scientific and industrial research activities, and in the functionalities of the two kinds of IT considered: BITNET versus Internet.

More broadly, while our analyses examine collaborations among researchers in locations within the United States, our results speak to research on the benefits and costs of geographically dispersed

[2] For examples of this work, see Agrawal and Goldfarb (2008), Ding et al. (2010), Rosenblat and Mobius (2004), Winkler, Levin, and Stephan (Forthcoming), or Walsh and Bayma (2006). For an example of work that examines theoretically the role IT can play in linking dispersed communities, see van Alstyne and Brynjolfsson (2005).

collaborations that has usually been conducted on samples of multinational companies. As is well known, while geographically dispersed research organizations may be effective at assimilating local knowledge from outside of the firm (e.g., Kogut and Zander 1992; Frost et al. 2002), cross-regional transfer of knowledge is difficult and costly even within the boundaries of the firm (Teece 1977; Kogut and Zander 2003; Singh 2008; Sorenson et al. 2006). As a result, the evidence on whether geographic dispersion improves a firm's innovative capabilities remains mixed (e.g., Furman et al. 2006; Leiponen and Helfat Forthcoming). It is well known that collaborative work is a powerful enabler of knowledge transfers, however (e.g., Singh 2005; Fleming et al. 2007). By suggesting a beneficial effect of Internet adoption on distant collaborations, our paper is therefore in the spirit of recent work that has examined the implications of the use of coordinating mechanisms within firms to facilitate integration of knowledge across units (e.g., Singh 2008; Argyres and Silverman 2004).

Our research also speaks to work on the value of information technologies investments. Recent work has demonstrated that increases in IT spending are correlated with growth in intangible assets such as patents or trademarks at the firm level (e.g., Gao and Hitt 2004; Kleis et al. Forthcoming). However, none of these speak to the impact of IT investments on coordination costs or to changes in the patterns of research collaborations within firms, as we do.

Our research has important public policy implications. It has been argued for some time now that, by lowering costs of communication, increasing use of IT will facilitate the globalization of economic activity and in particular research activity (e.g., Cairncross 1997; Friedman 2005). There is increasing interest in measuring whether such IT investments have in fact facilitated increasing dispersion of innovative activity (e.g., Macher and Mowery 2008). However, as yet there is little evidence on the link between IT investments and the organization of research activity. This paper takes a first step toward presenting such evidence.

2. Motivation and Hypotheses

In this section we present a simple framework that will motivate a set of hypotheses on how a reduction in coordination costs enabled by investment in IT will lead to increases in inventive output among geographically dispersed research teams relative to other types of research collaborations. Our focus on inventive output rather than productivity reflects a data constraint: we do not possess project-level data on R&D expenditures for firms in our sample. Our research strategy will be to compare the impact of Internet adoption on research output among research teams where ex ante coordination costs are high (e.g., geographically dispersed collaborative teams) to those where ex ante costs are low (e.g., lone inventors). If we observe, for example, after adoption that research output among geographically dispersed teams increases while that of lone inventors falls, then our results will be informative about how the Internet influenced coordination costs and the relative productivity of different types of research collaborations.

Our framework and research design is motivated by Becker and Murphy's (1992) model of team formation (and Adams et al. (2005) adaptation to a research context) in that we view decisions about team composition as shaped by the division of labor, task specialization, and coordination costs. In these models, research output is determined by factors such as the number of collaborators, their skill level, and a productivity shifter. Increases in the number of collaborators will increase gross output through task specialization and division of labor. Further, if specialized skills are geographically dispersed throughout the firm, then research output may be increasing in the geographic dispersion of researchers.[3]

However, increases in team size and dispersion are also likely to increase coordination costs. In particular, cross-regional transfer of complex or tacit knowledge is known to be difficult, even within firm boundaries (e.g., Teece 1977; Kogut and Zander 1993; Sorenson et al. 2006). Further, concerns of free-riding and shirking may also be increasing in team size (e.g., Holmstrom 1982), and monitoring geographically dispersed team members may be particularly challenging.

By lowering communication costs, adoption of basic Internet can help to reduce coordination costs. For example, Internet technology can lower communication costs by providing access to Internet protocol (IP)-based email, telephony, and other collaborative tools (Rice 1994; Lee and Choi 2003). This will facilitate lower access cost to others, especially to researchers in distant locations who have relatively few alternative means of communication available. In short, adoption of basic Internet will lower coordination costs, particularly among geographically dispersed researchers.

We further note that adoption of basic Internet has the potential to influence research output in other ways than through lower coordination costs. For example, Internet technology facilitates access to codified knowledge (e.g., Ding, Levin, Stephan, and Winkler 2010) by lowering the costs of accessing shared resources such as electronic databases for journals and online repositories for data. It also facilitates the development of more efficient processes for accessing knowledge, as when an institution sets up an online mechanism for accessing books from a library. In short, adoption of basic Internet is likely to increase the total factor productivity for all types of research collaborations.

As a result of these declines to coordination costs and improvements to total factor productivity, adoption of basic Internet will lead to an increase in output from collaborative, geographically dispersed research teams.

Hypothesis 1a: Adoption of basic Internet will be associated with an increase in output from collaborative, geographically dispersed teams.

[3] For example, Adams et al. (2005) consider the case where the average skill level of researchers is increasing with geographic dispersion.

The implications of basic Internet adoption for output from other types of research teams are more ambiguous. We consider the impact of basic Internet on two alternative types of teams: the case of collaborative teams within a geographic location and the case of lone inventors. For both of these types of groups, coordination costs will fall by less than for geographically dispersed teams. Productivity for geographically dispersed teams will rise by more than for other types of research groups, leading to a potential shift in research inputs toward geographically dispersed teams: this shift in resources may lead to a decline in research output for collocated and lone inventor teams. However, as noted above total factor productivity for all types of teams may rise due to declines in the costs of accessing shared resources, so it is also possible that research output may increase for these groups.

In short, it is difficult to sign ex ante whether Internet adoption will lead to an increase or fall in output for single-location collaborative teams: the increase in total factor productivity from declines in shared resource access costs may be offset by a shift in resources toward multi-location collaborations. However, our framework does predict clearly that the increase in output for these groups will be lower than for geographically dispersed teams.

Hypothesis 1b: Adoption of basic Internet will be associated with a lower increase in output for single-location collaborative teams than for geographically dispersed teams.

Hypothesis 1c: Adoption of basic Internet will be associated with a lower increase in output for lone inventors than for geographically dispersed teams.

Where did Internet adoption most influence collaborative research output?

Our next set of hypotheses examines the conditions under which adoption of basic Internet will be associated with the largest increase in collaborative inventive output. We focus in particular on the implications for geographically dispersed teams, where ex ante coordination costs are highest and the implications for inventive output are most clear. We examine three comparative statics implications arising from ex ante differences in organizational research productivity, gains to specialization, and coordination costs.

We first examine how economies of scale and scope in the locations where researchers are located influence the comparative statics of a decline in coordination costs. Prior work has demonstrated that increases in the size of research operations can improve productivity through economies of scale and scope (Panzar and Willig 1981; Cohen and Levin 1989; Henderson and Cockburn 1996). For example, larger research operations can spread the substantial fixed costs of inputs such as large pieces of equipment over a large base of research activity (Henderson and Cockburn 1996). Further, research operations may benefit from economies of scope by sharing knowledge inputs across research activities and programs.[4] In our setting we are unable to identify between the effects of economies of scale and scope and instead treat those both as implications of larger research group size.

Thus, by improving the productivity of a firm's research efforts, economies of scale and scope will influence the comparative statics of a decline in coordination costs. Formally, economies of scale and scope can be thought of as increasing the total factor productivity of the research output function. Productivity gains arising from the adoption of basic Internet will be even larger in the presence of economies of scale and scope, leading to a greater increase in output.

Hypothesis 2: Adoption of basic Internet will be associated with a larger increase in output from collaborative, geographically dispersed teams when adopted in research locations with economies of scale and scope.

Next we examine how the comparative statics of a fall in coordination costs vary with the size of ex ante coordination costs. Because of challenges of measuring their ex ante size, we focus on one particular dimension of ex ante coordination costs: the extent of prior collaboration among researchers in the organizations adopting IT. One mechanism that has been known to facilitate the flow of knowledge has been strong interpersonal network ties across heterogeneous units (e.g., Kogut and Zander 2003; Hansen 1999; Singh 2005, 2008; Sorenson et al. 2006). Such ties have the potential to influence coordination costs as they may act as knowledge brokers to integrate ideas from different regions (Burt 2004). Further, individuals who are unfamiliar with one another may be less willing to help one another due to concerns of opportunism. Such concerns could hinder within-organization transfer and integration of knowledge and increase coordination costs.

Adoption of IT may be particularly effective in such environments. Prior research has shown that in certain circumstances technological tools such as knowledge repositories can be used to facilitate knowledge transfer (e.g., Argote and Ingram 2000). Thus, adoption of Internet technology may be particularly effective at lowering coordination costs in environments without prior collaborations because of its ability to assist in the within-organization transfer and integration of knowledge.[5]

Hypothesis 3: Adoption of basic Internet will be associated with a larger increase in output from collaborative, geographically dispersed teams when adopted in research locations that have not collaborated before.

---

[4] For further details on the microfoundations of how such knowledge exchange might take place, see Singh and Fleming (2010) and Fleming et al. (2007).

[5] We acknowledge that lack of prior collaboration may also signal variance in the benefits of collaboration: research organizations that have not collaborated before may simply have few research synergies and few collaboration opportunities. Under these conditions, adoption of IT will have a weaker impact on collaborative output among research organizations. We leave this as an empirical question. If lack of prior collaboration does signal fewer collaboration opportunities, then it will bias our estimates of the effects of higher coordination costs downward.

The models of team collaboration that motivate our hypotheses predict that the marginal product of increases in team size will be greatest when the knowledge embodied in the human capital of workers is highly specialized (Becker and Murphy 1992). In these environments the benefits of division of labor are particularly large; an individual worker is unable to acquire the entire knowledge base necessary for production except at great cost. Several authors have argued this result has contributed to a long run trend of increases in team size observed in knowledge work like research (Becker and Murphy 1992; Jones 2009): For example, the increasing burden of knowledge from technological advances in science has led to increasing incentives for researchers to specialize, which has in turn lead to increases in collaborative research (Jones 2009).

These models have two implications for which types of research collaborations will see the strongest increase in output from Internet adoption. First, at any point in time there is significant variance in the burden of knowledge and value of specialization across scientific fields. Thus, there will be significant variance in the marginal product of an additional team member across fields. Thus, the implications of a fall in coordination costs that facilitates larger team sizes will vary significantly across scientific fields. Research teams that work on fields for which the value of specialization and division of labor is particularly high will see the greatest increase in output from Internet adoption.

Hypothesis 4a: Adoption of basic Internet will be associated with a larger increase in output from collaborative, geographically dispersed teams when adopted in research locations that work in research fields that require greater specialization.

Second, these models suggest that the benefits of a fall in coordination costs will be particularly large when Internet has been adopted in locations where researchers have specialized in heterogeneous fields. In such environments, research groups will be able to immediately capitalize upon the benefits of task specialization and division of labor. The concentration (or dispersion) of research fields reflects the homogeneity (or diversity) of skills in the corresponding entity and therefore expresses the degree of task specialization inside it. To be clear, this hypothesis and the previous one both capture variance in the benefits of Internet adoption (for geographically dispersed teams) based upon the extent to which the adopting location will benefit from additional task specialization. While the prior hypothesis uses a measure of task specialization based upon the type of research field, this one measures specialization through the dispersion across fields in the location under observation.

Hypothesis 4b: Adoption of basic Internet will be associated with a larger increase in output from collaborative, geographically dispersed teams when adopted in research locations where research is spread more across technological fields.

3. Empirical Strategy

3.1 Adoption of Internet technology and collaborative output

We argue that adoption of basic Internet will be associated with a decline in coordination costs for research teams. As a result, we expect an increase in research output from collaborative, geographically dispersed teams. To examine whether the empirical evidence is consistent with this hypothesis, we seek to measure the impact of Internet adoption on multi-inventor collaborations in geographically dispersed firm-location pairs.

We use a difference-in-difference identification strategy, comparing the number of (citation-weighted) collaborative patents in a firm-location pair of a time period before basic Internet technology diffused (1992) to those of a period where we observe adoption (1998).[6] This approach allows us to remove unobserved firm-pair features that may be correlated with Internet adoption and patents. This yields the following regression equation:

$$\log(Patents_{ijk1998}) - \log(Patents_{ijk1992}) = \alpha_1 X_{ijk} + \alpha_2 Z_{ijk} + \beta Internet_{ijk} + \varepsilon_{ijk} \quad (1)$$

The variable $Internet_{ijk}$ measures whether both establishments *j* and *k* in the pair of a particular firm *i* adopted basic Internet at time *t*. Internet technology had not diffused among firms prior to 1995 except in very rare cases, so we set the value of this variable to zero in 1992. We have two types of controls: the variables in X_{ijk} capture changes in firm-pair controls for things like (the log of per-establishment) firm R&D expenditures and firm-location employment that may affect the volume of collaborations in a firm-pair. The variables in Z_{ijk} capture changes in local characteristics that may influence inventive output. We have assumed that ε_{ijk} is a normal i.i.d. variable, but use robust standard errors in our estimation.

As noted above, our endogenous variable is $\log(1+Patents_{ijkt})$, which represents the number of citation-weighted patents applied for in year *t* with inventors in both locations within the pair of firm *i*. Citation-weighted patents have been used extensively as a measure of inventive output, however there are, of course, significant limitations to their use in this way. As Jaffe and Trajtenberg (2002) note, not all inventions meet the U.S. Patent and Trademark Office (USPTO) criteria for patentability. Further, inventors must make an explicit decision to patent an invention, as opposed to relying on some other method for intellectual property protection. In particular, there may be incremental inventive activity that is not patented and therefore is not reflected in patent statistics (e.g., Cohen, Nelson, and Walsh 2000). Firms may sometimes also choose to use trade secrecy rather than patenting to protect groundbreaking inventions because of incomplete enforcement of property rights. However, citation-weighted patents have been shown to be correlated with a firm's stock market value, and thereby provide one useful measure of a firm's intangible stock of knowledge (Hall, Jaffe, and Trajtenberg 2005). Further, so long as

[6] The results are robust to the use of 1990, 1991 and 1993 instead of 1992 as reference years. Further details on these results are included in the Appendix.

a firm-location's patent propensity does not vary significantly over time in a way that is correlated with Internet adoption, this should not bias our estimates of the key parameters of interest.

Our hypothesis is that the adoption of basic Internet at both locations in the firm-pair will be associated with an increase in the number of collaborative inventions, as proxied by the number of (citation-weighted) collaborative patents: a test of β>0 against the null of β=0.

As noted above, Internet adoption may also be correlated with an increase in collaborative output within firm-location pairs. However, we expect the relationship to be weaker because the decline in coordination costs will be lower than in the cross-location case (Hypothesis 1b). To measure the impact of basic Internet adoption on within-location collaborations, we estimate a variant of the above equation for collaborations within a single MSA. Our endogenous variable will be log($Patents_{ijt}$), which represents the number of patents applied for in year *t* with at least two inventors in location *j* of a particular firm *i*.

$$\log(Patents_{ij1998}) - \log(Patents_{ij1992}) = \alpha_1^{sl} X_{ij} + \alpha_2^{sl} Z_{ij} + \beta^{sl} Internet_{ij} + \varepsilon_{ij} \quad (2)$$

Here, $Internet_{ij}$ is a binary indicator of whether basic Internet has been adopted at the location, and X_{ij} and Z_{ij} represent changes in firm-location and location level controls, respectively. log($Patents_{ijt}$) represents the number of collaborative citation-weighted patents applied for in year *t* with inventors located only in location *j* of firm *i*: this measure includes only patents with multiple inventors, all of whom are located in a single location. As noted above, we expect the marginal effect of basic Internet adoption on the rate of growth in patenting for collocated inventors to be smaller than for geographically dispersed inventors. In fact, if the effect on coordination costs is small and if basic Internet adoption has little effect on the costs of shared resource use, then we may observe $\beta^{sl} = 0$.

Further, to examine whether basic Internet adoption is associated with an increase in single-inventor patents, we re-estimate equation (2) using only single-authored patents (*SAPatents*)

$$\log(SAPatents_{ij1998}) - \log(SAPatents_{ij1992}) = \alpha_1^{slsa} X_{ij} + \alpha_2^{slsa} Z_{ij} + \beta^{slsa} Internet_{ij} + \varepsilon_{ij} \quad (3)$$

We expect the marginal effect of basic Internet adoption to be lower here than in the case of multiple inventors, as there will be no effect on coordination costs. In fact, if the adoption of basic Internet has no effect on costs of shared resource usage, then we may observe $\beta^{slsa} = 0$ or even $\beta^{slsa} < 0$.

We initially assume that there are no unobserved factors in ε_{it} in any of equations (1) through (3) that are correlated with basic Internet adoption. We then explore this assumption. A particular concern is omitted variable bias, whether changes in unobserved features of the firm-location pair may be correlated both with Internet adoption and collaborative patent growth. We do several things to explore this assumption. First, we perform several sets of analyses to circumscribe how unobserved factors may influence our results. We conduct a falsification exercise where we examine if the number of collaborative patents over a period (1990-1994) prior to the diffusion of the commercial Internet is correlated with an establishment's later adoption of Internet technology (i.e., in 1998). We further examine, within the context of the regression described in equation (1), whether Internet adoption at one location in the pair is associated with an increase in patent output. In particular, if Internet adoption is associated with an increase in cross-location collaborative patenting due to a decline in coordination costs, then we should observe no impact on patenting when only one location in the pair adopts Internet technology. This is exactly what we find, providing additional evidence in support of our theory.

We also demonstrate that our results are robust to the use of instrumental variables. One instrument we employ—the average year of price cap regulation in the states in which Internet is adopted—proxies for local telecommunications costs. Another instrument—the penetration of basic Internet into the establishment's 3-digit NAICS industry—captures differences in average industry-level benefits to adoption. Further details on these instruments are discussed below.

3.2 Measuring variance in the effects of Internet adoption

To measure variance in the impact of basic Internet adoption across different environments, we interact our Internet adoption variable with proxies for economies of scale and scope, prior collaborations, and importance of specialization and division of labor. To capture the effects of economies of scale and scope, we compute $HighPriorPatents_{ijk}$, an indicator of whether the sum of citation-weighted patents in the pair over 1990-1992 is in the top quartile of our sample. As a robustness check, we follow Henderson and Cockburn (1996) in using R&D spending as a measure of economies of scale and the results are qualitatively similar.[7] To capture the effects of prior collaborations, we compute $NoPriorCollab_{ijk}$, which is equal to one if the pair had no prior collaborations over the period 1990-1992. As noted earlier, one way of measuring ties is by calculating past joint projects as proxied by research papers or patents (e.g., Cockburn and Henderson 1998). Last, to measure the importance of specialization we compute $HighCollabClass_{ijk}$. To compute this variable, we first identify the 10 research categories in which the number of inventors per patent was the largest on average over the period 1990-1992, among the 36-categories identified by Hall et al. (2001). Our *HighCollabClass* variable is equal to 1 if one of the locations in the pair is in the top quartile of firm-location pairs with the largest shares of patents in these 10 highly collaborative classes.[8] Jones (2009) shows that there is a positive

---

[7] As elsewhere, we compute firm-level R&D spending and then deflate it by the number of firm locations to obtain a per-location measure of scale. For brevity, these results were not included in the paper but are available from the authors upon request.

[8] We have experimented with alternative thresholds for this variable (e.g., the top 5 most collaborative classes) and the qualitative results remain similar.

correlation between team size and the burden of knowledge for a field, measured by size of the citation tree behind any patent. Thus, $HighCollabClass_{ijk}$ is a measure of the benefits of specialization and division of labor. To capture the effects of field specialization across locations in the pair, we compute the variable $HighSpecialization_{ijk}$. To construct this variable, we compute the Herfindahl index of the distribution of patents, across both locations in the pair, using the Hall-Jaffe-Trajtenberg (2001) six field categories. We set $HighSpecialization_{ijk}$ equal to one when this index is in the top quartile of the distribution.[9]

To test hypotheses 2 through 4, we add interactions of each of these variables to equation (1).

$$\begin{aligned}\log(Patents_{ijk1998}) - \log\left(Patents_{ijk1992}\right) \\ = \alpha_1 X_{ijk} + \alpha_2 Z_{ijk} + \beta_1 Internet_{ijk} + \beta_2 Internet_{ijk} \times HighPriorPatents_{ijk} \\ +\beta_3 Internet_{ijk} \times NoPriorCollab_{ijk} \\ +\beta_4 Internet_{ijk} \times HighCollabClass_{ijk} \\ +\beta_5 Internet_{ijk} \times HighSpecialization_{ijk} + \varepsilon_{ijk}\end{aligned}$$

A test of Hypothesis 2 represents a test of $\beta_2 > 0$ compared to the null of $\beta_2 = 0$; a test of Hypothesis 3 represents a test of $\beta_3 > 0$ against a null of $\beta_3 = 0$; a test of Hypothesis 4a represents a test of $\beta_4 > 0$ against a null of $\beta_4 = 0$; and a test of Hypothesis 4b represents a test of $\beta_5 > 0$ against a null of $\beta_5 = 0$

4. Data

We use a variety of data sources to show how adoption of basic Internet influences collaborative research output within firms. In particular, we match data on IT investment from a well-known private data source on IT expenditures with patenting data from the USPTO. We first describe our patent data, then our IT investment data, and then the construction of control variables. Last we describe the construction of our pairs data set.

Patent Data. We use patent data from the US Patent and Trademark Office (USPTO) as a measure of inventive activity. Patents are dated using the year of application because of the variance in the patent application-grant delay over time, and because application dates are closer to the time when the innovation occurred (e.g., Griliches 1990). Because of the well-known heterogeneity in the value of patents, we weight patents by citations using the procedure described in Hall, Jaffe, and Trajtenberg (2001). Further, we consider only citations within five years of the grant to avoid truncation bias (van Zeebroeck 2011), and deflate the citations received by each patent by its IPC4-year average to control for

[9] Again, we experimented with alternative ways of constructing this variable (like using the continuous Herfindahl measure) and the results were qualitatively similar.

cross-industry differences in the propensity to patent and cite other patents (Hall, Jaffe, and Trajtenberg 2001).

We map patents to firm identifiers using the patent's assignee information and the NBER Patent Data Project's matching data set which maps patents to a consistent set of unique firm identifiers based on the "GVKEY" code from the COMPUSTAT database. We obtain the universe of patents with a matching GVKEY that were applied for during 1990-1998.

Our analyses will examine the geographic variance in patenting behavior across firm-Metropolitan Statistical Areas (MSAs).[10] Using the inventor location data in US patents, we map inventors to MSAs using the zip code of the inventor (obtained through the USPTO Patents BIB data product). In cases where Consolidated Metropolitan Statistical Areas (CMSAs) were present, we used those, because it better allowed us to capture commuting patterns.[11] In regions of the US that are outside of MSAs, we constructed "phantom" MSAs that consisted of the region of a state outside of all of the MSAs. Our procedure will accurately map patents to the MSA they were invented in, to the extent that inventors work in the same MSA where they reside. MSAs are constructed in part on the basis of commuting patterns and are widely used as a unit of analysis in studies of the geography of innovation (e.g., Feldman and Audretsch 1999), however our procedure may inaccurately assign some patents to the wrong MSA when one or more of its inventors commutes to or from a different MSA.

IT Data. Our data on IT investment come from the Harte Hanks Market Intelligence Computer Intelligence Technology database (hereafter CI database). The database contains establishment- and firm-level data on characteristics such as the number of employees, personal computers per employee, and use of Internet applications. Harte Hanks collects this information to resell as a tool for the marketing divisions of technology companies. A number of researchers have used this data previously to study adoption of IT (e.g., Bresnahan and Greenstein 1996) and the productivity implications of IT investment (e.g., Brynjolfsson and Hitt 2003; Bloom et al. 2009). Interview teams survey establishments throughout the calendar year; our sample contains the most current information as of December 1998.

Harte Hanks tracks over 300,000 establishments in the United States. Because we focus on industrial research, we exclude government, military, and nonprofit establishments. Our sample from the CI database contains commercial establishments with over 100 employees. While this limits our sample

[10] This choice is made in part due to a data constraint. While our IT data are in fact available for individual firm establishments, USPTO patent data provide only inventor locations. Thus, for multi-establishments MSAs, we are unable to identify the particular establishment at which an inventor works within an MSA.

[11] CMSAs represent regions that may contain multiple metropolitan areas, such as Baltimore, MD-Washington, DC or San Francisco-Oakland-San Jose. We have rerun our statistical analyses using these component areas (Primary Metropolitan Statistical Areas, or PMSAs) instead of CMSAs and while the results are qualitatively similar, they are somewhat weaker. We attribute these weaker results to measurement error induced by inaccurate mapping of inventors to PMSAs due to commuting patterns of inventors across PMSAs within the same CMSA.

to predominately large establishments, our algorithm for matching our IT data to firms using Compustat identifiers from the NBER Patent Data Project similarly requires us to focus upon large firms. Further, our primary research question—how the adoption of the commercial Internet affected the geography of research collaborations within firms—also circumscribes our focus to large, multi-establishment research organizations. Thus, our analysis should be viewed as a study of IT and research collaborations within large research organizations. Prior work has compared the Harte Hanks data to the distribution of establishments in the Census County Business Patterns and found that the data include slightly less than half of all establishments with over 100 employees in the United States, and represents roughly one-third of all employment (Forman, Goldfarb, and Greenstein 2002).

Our raw data include at least twenty different specific applications, from basic access to software for Internet-enabled ERP business applications software. As noted earlier, we focus on that set of applications and technologies that involve little adaptation by users to be implemented successfully: these are typically some of the technologies that diffused around the initial commercialization of the Internet such as access to the Internet and the creation of static web pages within an organization. Our focus on this set of technologies reflects our interest in understanding how lower communication costs lowered the coordination costs of geographically dispersed, highly collaborative research.

We define an establishment as a basic Internet adopter if it indicates that it has one of the following in 1998: basic access to the Internet (i.e., whether the establishment has an ISP), an internal intranet based on TCP/IP protocols, or uses the Internet for research purposes.[12] In particular, we do not require establishments to adopt electronic commerce or TCP/IP-enabled business applications software. Our measure of Internet adoption is meant to capture whether the establishment has adopted enabling technology that will lower communication costs. We set the value of basic Internet equal to zero for all establishments in 1992 as this was well before the diffusion of the commercial Internet.[13] While our measure of basic Internet adoption shares some similarities with the measure of Internet participation used in earlier studies of Internet diffusion based on Harte Hanks data (e.g., Forman, Goldfarb, and Greenstein 2005), there are some differences. In particular, we focus on a narrower set of applications than Forman, Goldfarb, and Greenstein do because of our focus on an earlier time period (1998 versus 2000) and changes over time in the questions asked by Harte Hanks.

[12] An alternative measure of basic Internet use would incorporate the use of TCP/IP-based email, however over some periods of our data it is difficult to identify email based on Internet protocols from that which is based on proprietary networking protocols that were still commonly used over our sample period. To the extent that basic access is required for the use of Internet-based email, we believe our measure captures the use of such email in our sample.

[13] While it is difficult to date the rise of the commercial Internet, as a point of reference Netscape's browser became available in early 1995, followed by its IPO in December of the same year.

As noted above, CI data are collected at the establishment level. To map our establishment-level IT data to our patent data, we match establishments to firm-MSAs as we had done with the patent data. We first map the unique firm identifier used in the CI database to the GVKEY from the NBER Patent Data Project. We then assign establishments to MSAs using their zip code. For our analysis data set, we include only firm-MSA-year triplets that are from manufacturing firms (SIC 20-40) and that are in firm-MSAs with at least one patent in two separate years over the period 1992-1998. These restrictions are to retain only firm-organizations that perform research for our analyses (many CI database establishments perform no research function); our results are robust to alternative sample restrictions such as firm-MSAs with at least one patent over 1992-1998. In cases where there are multiple establishments within an MSA[14] we calculate a firm-location as adopting basic Internet when at least one has done so.

Firm-MSA pairs. The focus of our study is on the effects of IT investment on collaborative cross-location inventive output. We estimate the regression model in equation (1) which allows us to examine, for each pair of firm-MSA establishments, whether the adoption of basic Internet technology in both locations is associated with an increase in the number of collaborative patents invented by inventors located in both locations. To do this, we form the complete set of pairwise combinations of Firm-MSAs within a given organization. Based upon co-authorship, we identify the number of collaborations that were performed between units in different MSAs in a given patent-application year. We further use equations (2) and (3) to examine whether there is a relationship between basic Internet adoption and within-MSA output.

Other controls. We combine these data with additional information from a number of sources. The additional data are used to control for time-varying factors that may be correlated with basic Internet adoption and with patent output. First, to control for variance in R&D inputs across firms, we compute the flow of R&D spending dollars using COMPUSTAT and compute the per-location R&D flow dollars by normalizing total spending by the number of Firm-MSA locations in our data.[15] Second, we compute total firm-location employment as the sum of employment across establishments within the location. Unfortunately, our CI data begin at 1996 so we are unable to observe firm-location employment in 1992. We use 1996 employment values for this to observe some time trend in employment growth; all of our results are robust to removing the employment variable.

Next, we control for a number of local factors that may influence both the likelihood of basic Internet adoption as well as innovation productivity and the propensity to patent. The data sources for these measures are at the county level and are then matched to MSAs and computed for a Firm-MSA-year

[14] This is the case for 35% of the firm-MSAs in our analysis sample.

[15] An alternative procedure would be to deflate by the number of establishments. However, some establishments in our data do not engage in innovative activity. Further, since our output measure is based upon firm-location pairs, our procedure matches R&D input with innovative output.

triplet. For our cross-location pair regressions, these data items are then averaged across triplets in a pair.[16] We use the percent of manufacturing employment in the MSA, the average weekly wage in the MSA, and the log of MSA employment using US Census County Business Patterns data. Using the USPTO data, we also compute the log of the total number of patents in the MSA-application year. For the latter two (logged) measures, we compute the log of the average across the two MSAs in our pair regressions.

As noted above, we also control for firm-location employment using information from the CI database. In our pair regressions, we compute the log of the average employment across the two locations.

Descriptive statistics for all variables are provided in Table 1a for cross-MSA collaborations and in Table 1b for within-MSA collaborations. Correlations are reported in Tables 2a and 2b respectively.

5. Results

We first establish a relationship between the adoption of basic Internet and the number of collaborative patents at geographically dispersed research locations. We then show that there is no significant relationship between adoption and the number of collaborative patents invented by researchers within a location, nor between adoption and the number of single-inventor patents. We demonstrate that these results are robust to a variety of specifications and robustness checks, and to the use of instrumental variables. Last, we examine evidence on the conditions under which basic Internet adoption will have the greatest impact on collaborative inventive output. We show that basic Internet adoption has a particularly strong effect on cross-location research collaborations among pairs that had historically been high patenters (*HighPriorPatents*), among pairs that had not collaborated before (*NoPriorCollab*), among those that work on scientific fields that are highly collaborative (*HighCollabClass*), and among those who work on a dispersed set of research fields (low values of *HighSpecialization*).

5.1 Baseline Results

Tables 3a reports a non-parametric difference-in-difference analysis of citation-weighted patent counts for cross-MSA collaborations over 1992-1998 and according to their adoption (or non-adoption) of basic Internet. It suggests that a statistically and economically significant increase in collaborative patent output occurred for cross-location pairs adopting Internet over the period. MSA pairs that both adopted basic Internet had an average increase in the number of collaborative citation-weighted patents that was 0.13 patents higher than non-adopters over this period[17]; this compares to an average number of citation weighted patents of 0.18 for (future) adopters in 1992.

---

[16] For our analyses of patent output within a single MSA, the average value is equal to the value of the variable for the Firm-MSA-year triplet, since both triplets in the pair are equal to the same value.

[17] Non-adopters include pairs where neither and only one member of the pair adopted Internet.

In contrast, Tables 3b and 3c show that there is no significant difference between adopters and non-adopters of basic Internet in the change in number of citation-weighted patents within single MSAs over the same period, either for multi-inventor or single-inventor patents. In fact, for both the multi-inventor and single-inventor cases locations that adopted basic Internet experienced a slower growth in within-location patenting, although the results are not statistically significant.

In Table 4 we use the regression model in equation (1) to examine the implications of basic Internet adoption for collaborative patent output across firm-MSA pairs (Hypothesis 1a). Column 1 shows what we view as our baseline specification. If both establishments in the pair have basic Internet this translates into a 2% increase in the growth of the number of (citation-weighted) patents; these results are statistically significant at the 5% level. As we show below, this point estimate masks considerable heterogeneity on the impact of basic Internet on collaborative research productivity.

We explore further robustness in columns 2 through 7. Column 2 shows that our results hold when we use the level of patents rather than the log. In column 5 we examine whether our results are robust to the use of a balanced panel of data over the years 1992, 1994, 1996, and 1998. In this setting, the Internet adoption dummy is turned on as soon as basic Internet access is reported by both establishments for the first time, i.e. in 1994, 1996 or 1998. The results are robust. We focus primarily upon our difference-in-difference estimates throughout the rest of the paper for two reasons. First, establishments enter and exit from the CI database across years as a result of changes in the sampling strategy used by Harte Hanks; we focus on a simple difference-in-difference regression to eliminate risks of bias from sample selection. Second, the questions on Internet technology use change in subtle ways across years. In columns 6 and 7 we examine the robustness of our results to alternative stochastic assumptions. Column 6 shows the results of pooled Poisson QML regression estimates using the count of citation-weighted patents; column 7 includes NLS estimates in which we directly specify the conditional mean as an exponential function rather than assuming our data generating process is Poisson, this allows us to use our (non-integer) citation-weighted industry-deflated patent counts that are used in our linear models in columns 1-5. Our results are robust to each of these models. While we have experimented with QML Poisson models with conditional fixed effects, for many of our pairs the number of patents in both periods is equal to zero and so are dropped from the estimation sample. These conditional fixed effects results are qualitatively similar to our baseline model but not statistically significant. [18] We speculate that the weaker significance of these results reflect both the large number of observations dropped as well as our inability to deflate our dependent variable for industry-specific time trends in citation patterns.

[18] We have also estimated Poisson QML random effect regressions (as always, using cluster robust standard errors) and the results are qualitatively similar with a p-value of 0.105.

One potential concern with these estimates is that they may be affected by omitted variable bias. If there exist unobserved features related to a firm pair or its location that are changing over time in a way that is systematically correlated with basic Internet adoption and with collaborative patenting, then our parameter estimates for basic Internet adoption will be biased. Though we discuss our instrumental variable estimates below, here we describe two tests that help to circumscribe the way in which omitted variable bias may influence our estimates.

In column 4 we show the results of a falsification test that utilizes the timing of Internet adoption. As has been reported extensively elsewhere, the commercial Internet diffused rapidly beginning in 1995. Prior to that time, Internet access existed only in a few academic research institutions. If we observe an effect of Internet adoption on patenting behavior prior to 1995, then there exist serious concerns that our results may be influenced by omitted variable bias. If we only observe the “right” timing for our Internet variable, then this adds additional confidence that our results reflect a causal relationship. Column 4 shows that there is little impact on Internet adoption over the period 1990-1994: the coefficient on Internet adoption is small (0.0052) and insignificantly different from zero.

Second, following Agrawal and Goldfarb (2008), we examine whether basic Internet adoption at one firm location is correlated with the number of collaborative patents. If Internet adoption influences research productivity primarily by lowering coordination costs, then adoption at one location should have no impact on the growth in the number of patents. However, if basic Internet influences productivity by lowering the costs of accessing shared resources, then we may observe a relationship between single-location adoption and collaborative output. Column 3 shows that basic Internet adoption at one location has no impact on the growth in the number of collaborative patents invented by researchers in the pair. This result is consistent with the view that adoption of basic Internet influences collaborations by lowering coordination costs: we provide further evidence in support of this view in our tests of Hypotheses 1b and 1c. In terms of robustness, these results suggest that if omitted variable bias is influencing our results, it must do so only when both establishments adopt basic Internet. While it is possible that unobservables with these characteristics might exist, it is hard to identify what they might be.

In columns 1 through 6 of Table 5 we show the results of our model that explores the relationship between basic Internet adoption and within-location collaborative patents. The results suggest that there exists no correlation between basic Internet adoption and the growth in within-location collaborative patenting. In column 7 we examine the relationship between basic Internet adoption and single-authored patents. There is no statistically significant relationship between basic Internet adoption and single-authored patents (note that to save space we have not conducted the full set of robustness checks that we employ for our multi-inventor analyses; however, we have conducted analogs to the analyses in columns

2 through 6 of Table 5 and in all cases the effects of basic Internet adoption remain statistically insignificant). These results are consistent with Hypotheses 1b and 1c.

In sum, Tables 3 through 5 show that adoption of basic Internet was associated with an increase in collaborative geographically dispersed inventive output. However, there is no evidence of an increase in either collaborative output within a geographic location or in output from lone inventors. This evidence—together with the results on single-location adoption in column 3 of Table 4—show that while there exists evidence that basic Internet lowered coordination costs among researchers, there is little evidence that basic Internet significantly improved researcher productivity through access to shared resources, at least in our setting and over this specific time period.

In the Appendix we include the results of a variety of additional robustness checks for our collaborative cross-location pair results. Our results are robust to using 1991 or 1993 as the base year in our difference-in-difference specification. They are robust to using an 8-year window for citation-weighting. We experimented with dropping IT-producing industries from our sample because the relationship between IT investment and collaborative patent output may be different from these industries; our results are robust to these changes.[19] We further studied the robustness of our results to the use of patents that were not citation-weighted; our results were also robust to this change although some significance was lost. This latter result is unsurprising given the well-known skewness in the distribution of patent value.

5.1.1 Instrumental Variable Estimates of Baseline Results

To further address concerns about omitted variable bias, in Tables 6a and 6b we include the results of instrumental variable estimates.[20] Our first instrument proxies for local deployment costs. We use the year in which the local state capped prices that incumbent local exchange carriers (ILECs) could charge entrants.[21] Because it captures variance in local telecommunications regulation, this variable should be correlated with Internet adoption. However, since it is an exogenous governmental policy shock, it is unlikely to be correlated with inventive output. Our second instrument captures variation in the benefits to Internet adoption across industries. We compute the average adoption rate among establishments in the same 3-digit NAICS industry, excluding establishments in the focal firm. Because it captures variance in the value to other related firms from adopting Internet, it is unlikely to be correlated

[19] We use the definition of IT-producing industries in Jorgenson, Ho, and Stiroh (2005).
[20] We note that one particular source of omitted variable bias that may be a concern is if managers of the firm emphasize globalization of research in the organization, and use Internet adoption as a signal of their commitment to global research. We note that to the extent that our instruments are very likely to be uncorrelated with these changes in managerial focus, our use of instrumental variables should help to address this concern.
[21] We thank Avi Goldfarb and Shane Greenstein for providing this instrument to us.

with patenting. For all instruments we compute the instrument for each location in the pair and then take the average.

Table 6a presents our second stage results (Columns 1 through 3 use 2SLS while column 4 are LIML instrumental variable estimates), while Table 6b presents the first stage results of the 2SLS estimates. The first stage results in table 6b shows that the likelihood of internet adoption is increasing in the time to new regulation and in industry propensity to adopt basic Internet. The F-statistic for the first stage instruments ranges from 39.75 for our just-identified results using industry propensity to 146.02 for our just-identified results using the regulatory change; in all cases the test statistics are significant at the 1% level.

Column 1 of Table 6a shows our second stage results with our full set of instruments; the effects of Internet adoption on the growth of collaborative patent output remains statistically significant. An overidentification test on these estimates does not reject the null hypothesis that our instruments are orthogonal to the 2nd stage residuals (x^2=3.52814, p-value 0.3171). We also present estimates using different sets of instruments in columns 2 to 4. Column 2 presents a set of results using only our instrument that captures cross-industry benefits to adoption. Column 3 presents the results of estimates using only our price cap instrument. Because it is determined by state government legislative processes, it has the strongest case for exogeneity. Column 4 presents the results of LIML estimates using both instruments. The second stage results in all of these regressions continue to show that the effects of basic Internet adoption are again economically and statistically significant at the 10% level or above. While the coefficient estimates for Internet adoption in all four columns of Table 6a are greater than those in column 1 of Table 5, a Hausman test retains the null hypothesis that they are the same in all cases. In short, our instruments suggest a statistically significant relationship between Internet investment and collaborative patent output that is geographically dispersed.

5.2 Where were the effects of Internet on research collaborations strongest?

In this section we examine when the adoption of basic Internet was associated with the strongest growth in patenting among inventors in dispersed locations. In particular, we show that basic Internet has a particularly strong effect on cross-location research collaborations among pairs that had historically been high patenters (*HighPriorPatents*), among pairs that had not collaborated before (*NoPriorCollab*), among those active in research areas that are among the most highly collaborative (*HighCollabClass*), and among those who worked in a dispersed set of research fields (*HighSpecialization*).[22] As described

[22] We tested the extent to which a variety of other moderating factors might influence the marginal effects of Internet adoption. In particular, we examined whether the marginal effect of Internet was different for pairs that were geographically dispersed and also studied whether our results were different for locations with unequal

above, we compute each of these measures based upon the distribution of patenting behavior over 1990-1992. To reduce the extent of unobserved heterogeneity in our sample, we drop firm-pairs that include locations with no patents over this period.

Column 1 of Table 7 replicates the results in column 1 of Table 4 using only establishments with patents over the period 1990-1992. Over this sample our original results remain qualitatively similar, however less statistically significant because of the smaller sample size (we lose over one quarter of our observations) and lower power of the test (p-value 0.134). However, this average effect obscures considerable heterogeneity within our sample.

Column 2 shows how the results differ for pairs that were in the top quartile of patenting over 1990-1992. The results show that firm locations that were in the top quartile of patenting who adopt basic Internet experience a 3.1% faster rate of patenting growth than other pairs. In contrast, those who were not in the top quartile experience no additional growth in patenting from adoption. These results are consistent with the interpretation offered in section 2 that adoption of basic Internet will have the strongest impact among pairs of locations that already exhibit some economies of scale in their research operations.[23]

Column 3 compares how the effects of basic Internet varies for firm-location pairs who had and who had not collaborated during 1990-1992. Pairs who had not collaborated before and who adopt basic Internet have on average a 10.8% faster rate of growth in patenting than those without Internet (5% significance). In contrast, pairs who had collaborated before and who adopt basic Internet experience no additional growth in patenting. As noted above, this suggests that Internet adoption has the strongest effect among pairs whose coordination costs were ex ante high.

Column 4 examines whether the marginal effect of basic Internet adoption is different for firm-location pairs engaging in highly collaborative research areas. We identify such pairs as those with patents in the top quartile of the patenting classes with the largest number of per-patent inventors on average. We find that pairs active in highly collaborative patenting classes experience a 5.8% greater increase in patenting as a result of basic Internet adoption (significant at the 5% level), while those inactive in these classes experience no such gain.[24]

Column 5 examines how the marginal effect of adoption differs for firm-location pairs engaging in more focused technological areas, reflecting a lower diversity – hence specialization – in skills and

---

inventive output, as in Agrawal and Goldfarb (2008). In both cases, we were unable to reject the null hypothesis that distance and inequality of patent output had any moderating effect on Internet adoption.

[23] Our results are robust to substituting *HighPriorPatents* with per-establishment R&D over the pair.

[24] In fact, the estimation results suggest that such pairs experience a loss, however the results are not statistically significant.

labor. The table shows that the pairs that were ex ante working in more focused research areas had a 4.4% lower increase in patenting as a result of basic Internet adoption.

Column 6 reports the results of a specification with interactions for all four of *HighPriorPatents, NoPriorCollab, HighCollabClass*, and *HighSpecialization*. These results largely confirm the disproportionate effect of basic Internet adoption for pairs with each of these features that were identified in columns 2 through 5. Pairs that adopted basic Internet but were not in any of these categories in 1990-1992 experienced a negative impact from basic Internet adoption (significant at the 5% level); while we cannot observe research inputs in our data, we interpret these results as reflecting a shift in resources toward pairs with *HighPriorPatents, NoPriorCollab, HighCollabClass*, and *HighSpecialization* and away from pairs without these features.

To address concerns about omitted variable bias, Table 8 provides instrumental variable estimates for the results in of Table 7. We interact both of our original instruments—time to first price cap and average industry propensity to adopt basic Internet—with our binary variables measuring pair heterogeneity (*HighPriorPatents*, *NoPriorCollab*, *HighCollabClass*, and *HighSpecialization*). While for brevity we do not include the first stage results here, the F-statistic for the excluded instruments in all first stage results are significant at the 5% level or above.[25] The resulting estimates are qualitatively similar to those in Table 7. While the results for *HighPriorPatents* and for *HighCollabClass* are not significant at conventional levels, they are significant when included in the combined regression shown in Column 6. Again, Hausman tests retain the null that the coefficients in Table 8 and Table 7 are of similar magnitude.

6. Conclusion

We examine the implications of basic Internet adoption for reducing the coordination costs of industrial research teams. We match local (MSA) business IT investment data with local firm patenting activity and, using a difference-in-difference econometric estimation approach, find robust empirical evidence that basic Internet adoption is associated with increased growth of citation-weighted collaborative patents in geographically dispersed firm teams. On the contrary, we find no evidence of such a link between Internet adoption and within-location collaborative patents, nor do we find any evidence of a relationship between basic Internet and single-inventor patents. We interpret these results as evidence that basic Internet adoption lowered the coordination costs of larger, geographically dispersed research teams, however we find little evidence that basic Internet adoption was associated with increased research output through easier access to electronic knowledge systems or shared resources (at least over our sample period). We further find that the link between basic Internet adoption and cross-location patenting is greatest for firm pairs that had previously been larger patenters, had not collaborated before,

[25] The first stage results are available from the authors upon request.

which focused on the top collaborative research areas, and which worked in widely dispersed sets of research areas. The latter two sets of results provide some evidence that basic Internet use facilitated the specialization and division of labor that other researchers have highlighted as a long run trend in science (Jones 2009).

Our results stand in contrast to recent work on IT and academic research that has found that IT adoption leads to a disproportionately greater increase in collaborations among researchers who are geographically close to one another (Agrawal and Goldfarb 2008). There are several potential reasons for this difference in results. First, Agrawal and Goldfarb study BITNET, a predecessor network to the Internet. While the latter allows for content-rich information and knowledge exchanges, one of the main benefits of the former was to share scarce computing resources. Next, whereas we look at patented output, they focus on scholarly publications. The differences in costs and processes leading to these research outputs may also explain some of the differences that we observe. Finally, we look at within-firm industry collaborations while Agrawal and Goldfarb examine academic collaborations across universities. Geographic proximity is commonly thought to facilitate the formation of new relationships. Once relationships are formed, communication among existing partners can be facilitated through electronic channels. This mechanism has led to the argument that IT and face-to-face communication are complements to one another (e.g., Gaspar and Glaeser 1998; Charlot and Duranton 2006). However, in our setting, partnerships among researchers are likely set by the research goals of managers within the firm so the benefits of geographic proximity to identifying research partners is less important than in an academic setting.

Our results have implications for the literature on knowledge diffusion within firms. Whereas evidence of the well-known stickiness of knowledge has been observed even across units within the same firm (e.g., Teece 1977; Kogut and Zander 2003; Szulanski 1996), collaborative ties have been found to be a very efficient way to transfer knowledge across branches, institutions, or industry boundaries (e.g., Singh 2005; Fleming et al. 2007). By providing robust evidence that IT investments can enable distant industrial R&D collaborations, and hence facilitate cross-unit integration through a decrease in coordination costs, the present study suggests that IT investments have the potential to reduce the well-known localization of knowledge flows.

There is an abundant body of research on the productivity of IT investments and more recently some work on the implications of IT investments for the growth in intangible assets like trademarks and patents (e.g., Gao and Hitt 2004; Kleis et al. Forthcoming). However, because this latter work has focused on IT capital spending using firm-level data, it has been unable to unpack how IT investments lead to growth in intangibles. Our paper provides evidence that IT investments influenced coordination costs, but little evidence of improving productivity by lowering costs of access to shared resources or distant knowledge. This result has important implications for the design of research organizations within firms. In this way, we add to recent work in the IT productivity literature (e.g., Bloom et al. 2007) on the implications of different margins of IT investment for business value and organizational design

While our study only relies on US data and on local capabilities, it has important implications for the study of the globalization of research. In designing their international R&D organization, firms are often thought to choose between a centralized organization that provides higher control but prevents access to local knowledge spillovers, or a geographically dispersed and decentralized structure which enables tapping into local knowledge resources but induces higher coordination costs and more difficult knowledge sharing across firm units (e.g., von Zedtwitz and Gassmann 2002). By suggesting that Internet adoption can reduce coordination costs across distant R&D workers, our results suggest that IT investments may substantially alter this organizational trade-off and render decentralized R&D models more attractive, hence encouraging a higher geographic dispersion of R&D activities within firms.

From a managerial perspective, our results suggest that IT can be used to integrate geographically dispersed operations, either obtained through acquisition or deliberately dispersed due to a need to access local knowledge resources or markets. More broadly, they have implications for the long run design of research organizations within firms. Our results suggest that firms that wish to disperse their research organizations to either capitalize on lower costs or on local capabilities can do so with the knowledge that these dispersed researchers can be linked through their IT investments.

While our data is some of the best available, it is limited to one sample over one time period, therefore restricting the potential generalization of our conclusions. This limitation might be overcome in the future with a larger sample of IT investments, perhaps obtained through Census Bureau microdata. Further, future work may seek to understand how IT investments influence research collaborations in cross-country data. Extension to the cross-country context could have particularly interesting implications, as coordination costs will be higher while simultaneously the division of labor among researchers may be quite different: For example, work by Zhao (2006) suggests that firms use their internal research organization to substitute for the weak appropriability regimes in some countries. In addition, our study paves the way for further research on the effect of or more advanced kinds of IT investments, such as those that facilitate social networking.

Further, as noted above, our results raise several questions about the implications of IT investments for knowledge flows within organizations. Future work should examine whether new collaboration patterns enabled by IT have mediated new knowledge flows within organizations. More broadly, future research should examine to what extent IT investments have reduced or increased the importance of traditional channels of knowledge transfer, such as spatial, social, and employment relationships. We hope that our paper will help stimulate future work in these important areas.

References

Adams, J. D., G. C. Black, J. R. Clemmons, and P. E. Stephan. 2005. Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981-1999. *Research Policy* 34(3): 259-285.

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

Argyres, N. S. and B. Silverman. 2004. R&D, organization structure, and the development of corporate technological knowledge. *Strategic Management Journal* 25(8-9): 929-958.

Becker, G. S. and K. M. Murphy. 1992. The Division of Labor, Coordination Costs, and Knowledge. *The Quarterly Journal of Economics* 107(4): 1137-1160.

Bloom, N., R. Sadun, and J. Van Reenen. 2007. Americans do IT Better: US Multinationals and the Productivity Miracle. CEP Discussion Paper No 788.

Bloom, N., L. Garicano, R. Sadun, and J. Van Reenen. 2009. The distinct effects of Information Technology and Communication Technology on Firm Organization. NBER Working Paper #14975.

Bresnahan, T. and S. Greenstein. 1996. Technical Progress and Co-invention in Computing and in the Uses of Computers. *Brookings Papers on Economic Activity, Microeconomics* 1996; 1-83.

Brynjolfsson, E. and L. Hitt. 2000. Beyond Computation: Information Technology, Organizational Transformation, and Business Performance. *Journal of Economic Perspectives* 14(4): 23-48.

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

Burt, R. S. 2004. Structural holes and good ideas. *American Journal of Sociology* 110(2): 349-399.

Cairncross, F. 1997. *The Death of Distance* Cambridge, MA: Harvard University Press.

Charlot, S. and G. Duranton. 2006. Cities and workplace communication: Some Quantitative French Evidence. *Urban Studies* 43(8): 1365-1394.

Cockburn, I. and R. Henderson. 1998. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. *Journal of Industrial Economics* 46(2): 157-182.

Cohen, W. and R. C. Levin. 1989. Empirical Studies of Innovation and Market Structure. In *Handbook of Industrial Organization, Volume 2*, eds. Schmalensee, R. and R. Willig, New York: North-Holland, pp. 1059-1108.

Cohen, W., R. R. Nelson, and J. P. Walsh. 2000. Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (Or Not) NBER Working Paper 7552.

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.

Fleming, L., D. Chen, and S. Mingo. 2007. Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly* 52(3): 443-475.

Forman, C., A. Goldfarb, and S. Greenstein. 2002. Digital Dispersion: An Industrial and Geographic Census of Commercial Internet Use. NBER Working Paper 9287.

Forman, C., A. Goldfarb, and S. Greenstein. 2005. How Did Location Affect the Adoption of the Commercial Internet? Global Village vs. Urban Density. *Journal of Urban Economics* 58(3): 389-420.

Friedman, T. 2005. *The World is Flat: A Brief History of the Twenty-First Century*. New York: Farrar, Straus, and Giroux.

Frost, T. S., J. Birkinshaw, and P. Ensign. 2002. Centers of excellence in multinational corporations. *Strategic Management Journal* 23(11): 997-1018.

Furman, J. L., M. K. Kyle, I. Cockburn, and R. M. Henderson. 2006. Public & Private Spillovers, Location, and the Productivity of Pharmaceutical Research. *Annales D'Economie et de Statistique* 79/80: 1-24.

Gao, G. and L. Hitt. 2004. IT and Product Variety: Evidence from Panel Data. *Proceedings of the 25th International Conference on Information Systems (ICIS), Washington, DC, 2004.*

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

Griliches, Z. 1990. Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature* 28(December): 1661-1707

Hall, B.H., A. B. Jaffe, and M. Trajtenberg. 2001. The NBER Patent Citation 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.

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.

Holmstrom, B. 1982. Moral hazard in teams. *Bell Journal of Economics* 13: 324-340.

Henderson, R. and I. Cockburn. 1994. Measuring Competence? Exploring Firm Effects in Pharmaceutical Research. *Strategic Management Journal* 15(S1): 63-84.

Henderson, R. and I. Cockburn. 1996. Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery. *The RAND Journal of Economics* 27(1): 32-59.

Hicks, D. and J. S. Katz. 1996. Science policy for a highly collaborative science system. *Science and Public Policy* 23: 39-44.

Jaffe, A. and M. Trajtenberg. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge: MIT Press.

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

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

Jorgenson, D. W., M. S. Ho, and K. J. Stiroh. 2005. *Productivity, Volume 3: Information Technology and the American Growth Resurgence.* Cambridge: MIT Press.

Kleis, L., P. Chwelos, R. Ramirez, and I. Cockburn. Information Technology and Intangible Output: The Impact of IT on Innovation Productivity. Forthcoming, *Information Systems Research*.

Kogut, B. and U. Zander. 1992. Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science* 3(3): 383-397.

Kogut, B. and U. Zander. 2003. Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies* 34(6): 516-529.

Lee, H., B. Choi. 2003. Knowledge management enablers, processes, and organizational performance: An integrative view and empirical examination. *Journal of Management Information Systems* 20 (1): 179-228.

Leiponen, A. and C. E. Helfat. Location, Decentralization, and Knowledge Sources for Innovation. Forthcoming, *Organization Science*.

Macher, J. and D. Mowery. 2008. *Innovation in Global Industries: U.S. Firms Competing in a New World*. Washington, DC: National Academies Press.

Panzar, J. and R. D. Willig. 1981. Economies of Scope. *American Economic Review* 71(2): 268-272.

Presser, S. 1980. Collaboration and the quality of research. *Social Studies of Science* 10(1): 95-101.

Rice, R. E. 1994. Relating electronic mail use and network structure to R&D work networks and performance. *Journal of Management Information Systems* 11(1): 9-29.

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

Sauer, R.D. 1988. Estimates of the returns to quality and coauthorship in economic academia. *The Journal of Political Economy* 96(4): 855-866.

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(1): 77-96.

Singh, J. and L. Fleming. 2010. Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science* 56(1): 41-56.

Sorenson, O., J. W. Rivkin, and L. Fleming. 2006. Complexity, networks, and knowledge flow. *Research Policy* 35(7): 994-1017.

Szulanski, G. 1996. Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm *Strategic Management Journal* 17(SI): 27-43.

Teece, D. J. 1977. Technology transfer by multinational firms: the resource cost of transferring technological know-how. *Economic Journal* 87: 242-261.

van Alstyne, M. and E. Brynjolfsson. 2005. Electronic Communities: Global Village or Cyber Balkans? *Management Science* 51(6): 851-868.

van Zeebroeck, N. 2011. The Puzzle of Patent Value Indicators. *Economics of Innovation and New Technology* 20(1): 33-62.

von Zedtwitz, M. and O. Gassmann. 2002. Market versus technology drive in R&D internationalization: four different patterns of managing research and development. *Research Policy* 31(4): 569-588.

Walsh, J. and T. Bayma. 1996. Computer networks and scientific work. *Social Studies of Science* 26(3): 661-703.

Winkler, A., S. Levin, and P. Stephan. The Diffusion of IT in Higher Education: Publishing Productivity of Academic Life Scientists. Forthcoming, *Economics of Innovation and New Technology*.

Wuchty, S., B. F. Jones, and B. Uzzi. 2007. The Increasing Dominance of Teams in the Production of Knowledge. *Science* 316(May 18): 1036-1039.

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

Table 1a – Descriptive Statistics for Pairs Including Different MSAs (as of 1998)

Variable	Mean	Standard Deviation	Minimum	Maximum	Number of Observations
Log of Weighted Citations	0.0907	0.3850	0	4.6790	4800
Basic Internet in both locations	0.7027	0.4571	0	1	4800
Log of per-establishment R&D spending	3.2323	1.4900	-0.1823	7.7295	4800
Log of establishment employees	7.6508	1.0959	5.2983	11.6315	4800
Share of local employment in manufacturing	0.1924	0.0634	0.0391	0.4861	4800
Local average weekly wages	615.2087	85.6592	391.015	848.3290	4800
Log of local employment	13.9104	0.9401	10.4167	15.7005	4800
Log of number of local patents	6.9150	1.2174	1.6094	9.1314	4800

Table 1b – Descriptive Statistics for Within MSA Analyses (as of 1998)

Variable	Mean	Standard Deviation	Minimum	Maximum	Number of Observations
Log of Weighted Citations	1.3065	1.4484	0	6.9157	1078
Basic Internet in both locations	0.8377	0.3689	0	1	1078
Log of per-establishment R&D spending	3.2187	1.5001	-0.9715	7.7295	1078
Log of establishment employees	7.3770	1.2132	5.2983	11.8936	1078
Share of local employment in manufacturing	0.1906	0.0810	0.0204	0.5483	1078
Local average weekly wages	621.5126	123.4558	390.9388	860.2807	1078
Log of local employment	13.6685	1.3018	10.2118	15.8465	1078
Log of number of local patents	6.5508	1.7094	0	9.1411	1078

Table 2a – Correlation Table for Pairs Including Different MSAs (as of 1998)

	Log Collaborative Citations	Basic Internet	Log R&D Spending	Log Establishment Employees	Share Manuf Employment	Average Weekly Wages	Log local employment	Log local patents
Log Collaborative Citations	1.0000							
Basic Internet	0.0435	1.0000						
Log R&D Spending	0.1371	0.0113	1.0000					
Log Establishment Employees	0.2311	0.1210	0.4375	1.0000				
Share Manuf Employment	-0.0379	-0.0362	-0.2239	-0.0418	1.0000			
Average Weekly Wages	0.1353	-0.0360	0.2407	0.1581	-0.5158	1.0000		
Log local employment	0.0858	-0.0119	0.1331	0.1069	-0.4144	0.7691	1.0000	
Log local patents	0.1100	-0.0112	0.1725	0.1066	-0.4248	0.8397	0.9222	1.0000

Table 2b – Correlation Table for Within MSA Analyses (as of 1998)

	Log Collaborative Citations	Basic Internet	Log R&D Spending	Log Establishment Employees	Share Manuf Employment	Average Weekly Wages	Log local employment	Log local patents
Log Collaborative Citations	1.0000							
Basic Internet	0.0723	1.0000						
Log R&D Spending	0.3273	0.0106	1.0000					
Log Establishment Employees	0.5010	0.1477	0.2651	1.0000				
Share Manuf Employment	-0.1151	0.0320	-0.1822	0.0248	1.0000			
Average Weekly Wages	0.2667	-0.0299	0.1826	0.0743	-0.5018	1.0000		
Log local employment	0.1785	-0.0114	0.1135	0.0289	-0.3847	0.7758	1.0000	
Log local patents	0.2454	-0.0010	0.1515	0.0413	-0.3702	0.8402	0.9193	1.0000

Table 3a – Citation Weighted Patents by Year and Whether Treated by Internet Adoption, Firm-MSA Pairs

	Before Treatment (1992)	After Treatment (1998)	First Difference (row)
Received Internet Treatment	0.1822 (N=3373)	0.3626 (N=3373)	0.1804** (N=3373)
Did Not Receive Internet Treatment	0.1256 (N=1427)	0.1766 (N=1427)	0.0509 (N=1427)
First Difference (column)	0.0566+ (N=4800)	0.1860** (N=4800)	*Difference in Difference* 0.1295* (N=4800)

We base this analysis on the sample of firm-location pairs that are observed before and after the treatment in our sample. ** indicates the difference is significant at the 1% level. * indicates that difference is significant at the 5% level. + indicates that difference is significant at the 10% level.

Table 3b – Citation Weighted Patents by Year and Whether Treated by Internet Adoption, Within-MSA Analyses

	Before Treatment (1992)	After Treatment (1998)	First Difference (row)
Received Internet Treatment	13.5399 (N=903)	17.1483 (N=903)	3.6084+ (N=903)
Did Not Receive Internet Treatment	5.5956 (N=175)	11.4583 (N=175)	5.8628+ (N=175)
First Difference (column)	7.9443* (N=1078)	5.6900 (N=1078)	*Difference in Difference* -2.2544 (N=1078)

We base this analysis on the sample of firm-locations that are observed before and after the treatment in our sample. * indicates that difference is significant at the 5% level. ** indicates the difference is significant at the 1% level. + indicates that difference is significant at the 10% level.

Table 3c – Citation Weighted Patents by Year and Whether Treated by Internet Adoption, Within-MSA Analyses for Single-Inventor Patents

	Before Treatment (1992)	After Treatment (1998)	First Difference (row)
Received Internet Treatment	4.1100 (N=903)	4.1384 (N=903)	0.0284 (N=903)
Did Not Receive Internet Treatment	1.4536 (N=175)	2.4708 (N=175)	1.0172 (N=175)
First Difference (column)	2.6564 (N=1078)	1.6676+ (N=1078)	*Difference in Difference* -0.9888 (N=1078)

We base this analysis on the sample of firm- s that are observed before and after the treatment in our sample. * indicates that difference is significant at the 5% level. ** indicates the difference is significant at the 1% level. + indicates that difference is significant at the 10% level.

Table 4 – Baseline Results – Different CMSAs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Levels	Includes Only One Adopter	Falsification Exercise, 1990-1994	Fixed Effects Balanced Panel, 1992-1998	Pooled Poisson QML	Poisson NLS (Exponential), with weighted patents
Basic Internet in both locations	0.0203	0.1232	0.0213	0.0052	0.0181	0.4968	1.4251
	(0.0103)*	(0.0558)*	(0.0108)*	(0.0101)	(0.0080)*	(0.2942)+	(0.4091)**
Log of per-establishment R&D spending	0.0359	0.0844	0.0360	0.0279	0.0315	0.1781	0.0520
	(0.0099)**	(0.0564)	(0.0099)**	(0.0104)**	(0.0080)**	(0.0676)**	(0.1068)
Log of establishment employees	-0.0414	-0.2289	-0.0415		-0.0333	0.7406	1.0274
	(0.0225)+	(0.1175)+	(0.0225)+		(0.0209)	(0.0712)**	(0.1893)**
Share of local employment in manufacturing	-0.0435	2.2173	-0.0428	0.1657	0.0936	-0.5664	-5.9580
	(0.4489)	(1.9722)	(0.4490)	(0.5769)	(0.3341)	(1.9125)	(4.5342)
Local average weekly wages	0.0006	0.0042	0.0006	0.0014	0.0004	0.0057	0.0019
	(0.0002)*	(0.0015)**	(0.0002)*	(0.0006)*	(0.0002)*	(0.0020)**	(0.0033)
Log of local employment	-0.1444	-1.2333	-0.1445	0.0227	-0.0793	-1.1756	-1.5166
	(0.1104)	(0.6791)+	(0.1104)	(0.0826)	(0.0842)	(0.2093)**	(0.2757)**
Log of number of local patents	0.0322	0.3678	0.0324	0.0450	0.0242	0.9585	1.5990
	(0.0322)	(0.1580)*	(0.0323)	(0.0341)	(0.0228)	(0.2138)**	(0.3745)**
Internet in Either Location			-0.0099				
			(0.0218)				
Observations	4800	4800	4800	4505	18860	9600	9600
R-squared (within)	0.01	0.01	0.01	0.01	0.01		
R-squared (total)	0.73	0.67	0.73	0.69	0.63	0.2654	0.0903
Number of Groups					4715	4800	4800

Columns (1) through (3) are first difference models where variables represent first differences between 1998 and 1992, column (4) is a first difference model between 1990 and 1994, column (5) is a balanced panel model with data every other year from 1992-1998 (variables represent levels rather than differences), and columns (6) and (7) are two-period Poisson panel data models. Within R-squared values are from difference models or from the "within" fixed effects estimator, while "total" R-square represent those from an equivalent panel model that includes the explanatory power of the fixed effects.

Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 5 – Baseline Results – Single CMSAs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Levels	Falsification Exercise, 1990-1994	Fixed Effects Balanced Panel, 1992-1998	Pooled Poisson QML	Poisson NLS (Exponential), with weighted patents	Single inventor patents
Basic Internet in both locations	0.0030	-2.3079	-0.0209	-0.0048	-0.3067	-0.3174	-0.0430
	(0.0885)	(3.8448)	(0.0799)	(0.0425)	(0.3762)	(0.4869)	(0.0618)
Log of per-establishment R&D spending	0.3881	7.7864	0.2972	0.3431	0.3527	0.0865	0.2160
	(0.0600)**	(1.6810)**	(0.0722)**	(0.0494)**	(0.0639)**	(0.1171)	(0.0429)**
Log of establishment employees	0.0536	3.1897		0.0492	0.7672	0.8331	0.0964
	(0.1405)	(2.3951)		(0.1004)	(0.0724)**	(0.1274)**	(0.1283)
Share of local employment in manufacturing	-0.6776	55.5844	1.1804	-0.3367	-2.3349	2.2011	0.8214
	(2.0218)	(38.6008)	(2.3903)	(1.7118)	(1.2144)+	(2.2331)	(1.3435)
Local average weekly wages	-0.0005	0.0745	-0.0004	-0.0003	-0.0009	-0.0016	-0.0009
	(0.0009)	(0.0341)*	(0.0018)	(0.0008)	(0.0013)	(0.0020)	(0.0007)
Log of local employment	-0.0975	19.7850	0.0249	0.1938	-0.8806	-0.8411	0.2046
	(0.5384)	(23.5306)	(0.4442)	(0.4259)	(0.1314)**	(0.0984)**	(0.4327)
Log of number of local patents	0.4365	10.5532	0.5586	0.4063	0.9959	1.0573	0.2140
	(0.1203)**	(4.2243)*	(0.1242)**	(0.0798)**	(0.1505)**	(0.1917)**	(0.0888)*
Constant	-0.2133	-13.3695	0.0517	-4.9627	1.6503		-0.0667
	(0.1513)	(5.2617)*	(0.1392)	(5.5011)	(1.4620)		(0.1191)
Observations	1078	1078	1065	4276	2156	2156	1078
R-squared (within)	0.06	0.06	0.04	0.04	0.63	0.5854	0.03
R-squared (total)	0.85	0.91	0.88	0.84			0.84
Number of Groups				1069	1078	1078	

Columns (1) and (2) are first difference models where variables represent first differences between 1998 and 1992, column (3) is a first difference model between 1990 and 1994, column (4) is a balanced panel model with data every other year from 1992-1998 (variables represent levels rather than differences), and columns (5) and (6) are two-period Poisson panel data models. Column (7) is a first difference model of the number of single-inventor patents between 1998 and 1992. Within R-squared values are from difference models or from the "within" fixed effects estimator, while "total" R-square represent those from an equivalent panel model that includes the explanatory power of the fixed effects.

Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 6a – Instrumental Variable Estimates – Second Stage

	(1)	(2)	(3)	(4)
	Regulation and industry adoption	Industry adoption	Regulation	All Instruments, LIML
Basic Internet in both locations	0.1625	0.1429	0.2326	0.1628
	(0.0538)**	(0.0624)*	(0.1242)+	(0.0539)**
Change in log of per-establishment R&D spending	0.0293	0.0302	0.0261	0.0293
	(0.0105)**	(0.0108)**	(0.0113)*	(0.0105)**
Change in log of establishment employees	-0.0420	-0.0419	-0.0422	-0.0420
	(0.0225)+	(0.0225)+	(0.0227)+	(0.0225)+
Change in the share of local employment in manufacturing	-0.0527	-0.0514	-0.0572	-0.0527
	(0.4546)	(0.4527)	(0.4639)	(0.4546)
Change in local average weekly wages	0.0006	0.0006	0.0006	0.0006
	(0.0003)*	(0.0003)*	(0.0003)*	(0.0003)*
Change in log of local employment	-0.1677	-0.1645	-0.1792	-0.1678
	(0.1116)	(0.1119)	(0.1128)	(0.1116)
Change in log of local patents	0.0297	0.0301	0.0285	0.0297
	(0.0326)	(0.0324)	(0.0334)	(0.0326)
Constant	-0.1620	-0.1492	-0.2080	-0.1622
	(0.0473)**	(0.0527)**	(0.0850)*	(0.0473)**
Observations	4800	4800	4800	4800

Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 6b – Instruments – First Stage

	(1)	(2)	(3)
	Both instruments	Industry adoption only	Price cap only
Change in log of per-establishment R&D spending	0.0074	0.0108	0.0430
	(0.0125)	(0.0126)	(0.0128)**
Change in log of establishment employees	-0.0088	0.0056	-0.0106
	(0.0200)	(0.0199)	(0.0205)
Change in the share of local employment in manufacturing	-0.1924	-0.7476	0.6226
	(0.6121)	(0.6095)	(0.6118)
Change in local average weekly wages	0.0001	-0.0001	0.0001
	(0.0003)	(0.0003)	(0.0003)
Change in log of local employment	-0.1445	0.1881	-0.1714
	(0.1404)	(0.1311)	(0.1427)
Change in log of local patents	-0.0008	-0.0307	0.0473
	(0.0392)	(0.0394)	(0.0391)
First Price Cap or Freeze	0.0230		0.0231
	(0.0036)**		(0.0037)**
Industry Propensity Instrument	1.3785	1.3814	
	(0.1125)**	(0.1143)**	
Constant	-2.4491	-0.3124	-1.4987
	(0.3400)**	(0.0879)**	(0.3432)**
Observations	4800	4800	4800
R-squared	0.05	0.02	0.03

Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%. First stage LIML results are identical to those in column (1).

Table 7 – Where Is the Effect of Internet Adoption Strongest?

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Prior Patents High	No Prior Collaborations	Top Collaborative Classes	Concentration across technological classes	Combined
Basic Internet in both locations	0.0213	-0.0005	-0.0664	0.0092	0.0323	-0.1140
	(0.0142)	(0.0122)	(0.0520)	(0.0149)	(0.0153)*	(0.0529)*
Basic Internet X High Prior Patenting		0.0306				0.0407
		(0.0129)*				(0.0113)**
Basic Internet X No Prior Collaborations in Pair			0.1029			0.1206
			(0.0510)*			(0.0511)*
Internet X Top Collaborative Classes				0.0561		0.0622
				(0.0246)*		(0.0247)*
Internet X High Specialization					-0.0426	-0.0388
					(0.0174)*	(0.0181)*
Change in log of per-establishment R&D spending	0.0430	0.0435	0.0419	0.0483	0.0448	0.0499
	(0.0130)**	(0.0130)**	(0.0128)**	(0.0134)**	(0.0130)**	(0.0133)**
Change in log of establishment employees	-0.0660	-0.0621	-0.0756	-0.0687	-0.0581	-0.0678
	(0.0353)+	(0.0353)+	(0.0351)*	(0.0353)+	(0.0354)	(0.0352)+
Change in the share of local employment in manufacturing	0.0768	-0.0046	0.1891	0.1468	-0.0568	0.0563
	(0.6513)	(0.6531)	(0.6555)	(0.6526)	(0.6495)	(0.6557)
Change in local average weekly wages	0.0007	0.0006	0.0008	0.0006	0.0006	0.0007
	(0.0003)+	(0.0003)+	(0.0003)*	(0.0003)+	(0.0003)+	(0.0003)+
Change in log of local employment	-0.1748	-0.1718	-0.1906	-0.1401	-0.1739	-0.1501
	(0.1554)	(0.1551)	(0.1557)	(0.1574)	(0.1549)	(0.1573)
Change in number of local patents	0.0608	0.0654	0.0576	0.0534	0.0611	0.0551
	(0.0448)	(0.0451)	(0.0451)	(0.0452)	(0.0448)	(0.0456)
Constant	-0.0878	-0.0837	-0.0982	-0.0880	-0.0840	-0.0915
	(0.0421)*	(0.0419)*	(0.0421)*	(0.0420)*	(0.0421)*	(0.0420)*
Observations	3459	3459	3459	3459	3459	3459
R-squared (within)	0.01	0.01	0.02	0.01	0.01	0.02
R-squared (total)	0.74	0.74	0.74	0.74	0.74	0.74

Within R-squared values are from difference models or from the "within" fixed effects estimator, while "total" R-square represent those from an equivalent panel model that includes the explanatory power of the fixed effects. Robust standard errors in parentheses + significant at 10%; * significant at 5%; ** significant at 1%

Table 8 – Where Is the Effect of Internet Adoption Strongest (2nd stage IV Estimates)?

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Prior Patents High	No Prior Collaborations	Top Collaborative Classes	Concentration across technological classes	Combined
Basic Internet in both locations	0.2187	0.1787	0.0295	0.2058	0.2092	-0.0724
	(0.0742)**	(0.0722)*	(0.0838)	(0.0739)**	(0.0733)**	(0.0855)
Basic Internet X High Prior Patenting		0.0220				0.0454
		(0.0157)				(0.0142)**
Basic Internet X No Prior Collaborations in Pair			0.1908			0.2058
			(0.0570)**			(0.0570)**
Internet X Top Collaborative Classes				0.0415		0.0552
				(0.0278)		(0.0278)*
Internet X High Specialization					-0.0540	-0.0512
					(0.0199)**	(0.0204)*
Change in log of per-establishment R&D spending	0.0290	0.0311	0.0289	0.0332	0.0330	0.0409
	(0.0143)*	(0.0143)*	(0.0139)*	(0.0149)*	(0.0142)*	(0.0145)**
Change in log of establishment employees	-0.0704	-0.0670	-0.0876	-0.0723	-0.0599	-0.0751
	(0.0354)*	(0.0355)+	(0.0351)*	(0.0354)*	(0.0355)+	(0.0353)*
Change in the share of local employment in manufacturing	0.1817	0.1103	0.3759	0.2315	0.0002	0.1496
	(0.6666)	(0.6668)	(0.6688)	(0.6664)	(0.6646)	(0.6664)
Change in local average weekly wages	0.0007	0.0007	0.0009	0.0007	0.0006	0.0008
	(0.0004)*	(0.0004)+	(0.0003)**	(0.0004)*	(0.0004)+	(0.0003)*
Change in log of local employment	-0.1740	-0.1719	-0.2035	-0.1483	-0.1729	-0.1664
	(0.1577)	(0.1569)	(0.1585)	(0.1603)	(0.1566)	(0.1594)
Change in number of local patents	0.0445	0.0498	0.0407	0.0394	0.0468	0.0448
	(0.0457)	(0.0461)	(0.0463)	(0.0461)	(0.0456)	(0.0468)
Constant	-0.2201	-0.2009	-0.2216	-0.2177	-0.1998	-0.1741
	(0.0653)**	(0.0632)**	(0.0632)**	(0.0648)**	(0.0648)**	(0.0608)**
Observations	3459	3459	3459	3459	3459	3459

Robust standard errors in parentheses + significant at 10%; * significant at 5%; ** significant at 1%

Appendix Table 1 – Robustness checks

	(1)	(2)	(3)	(4)	(5)
	Uses 1991 base year	Uses 1993 base year	Uses raw count of patents	Uses 8-year window for citation-weighting	Excludes IT-producing industries
Basic Internet in both locations	0.0190	0.0312	0.0143	0.0208	0.0212
	(0.0105)+	(0.0104)**	(0.0092)	(0.0104)*	(0.0100)*
Change in log of per-establishment R&D spending	0.0404	0.0328	0.0394	0.0345	0.0255
	(0.0113)**	(0.0110)**	(0.0089)**	(0.0100)**	(0.0096)**
Change in log of establishment employees	-0.0239	-0.0133	-0.0238	-0.0440	-0.0290
	(0.0218)	(0.0238)	(0.0181)	(0.0224)*	(0.0096)
Change in the share of local employment in manufacturing	0.1894	0.0172	-0.4114	-0.1241	-0.4804
	(0.4364)	(0.4243)	(0.3458)	(0.4218)	(0.4601)
Change in local average weekly wages	0.0007	0.0004	0.0006	0.0007	0.0004
	(0.0002)**	(0.0002)+	(0.0002)**	(0.0002)**	(0.0002)
Change in log of local employment	0.0083	-0.1213	-0.1454	-0.1538	-0.2160
	(0.1087)	(0.1080)	(0.0939)	(0.1067)	(0.1058)*
Change in number of local patents	0.0087	0.0436	0.0375	0.0288	0.0092
	(0.0315)	(0.0311)	(0.0271)	(0.0328)	(0.0262)
Constant	-0.0827	-0.0591	-0.0702	-0.0738	-0.0278
	(0.0296)**	(0.0294)*	(0.0273)*	(0.0293)*	(0.0276)
Observations	4800	4800	4800	4800	4335
R-squared	0.01	0.01	0.01	0.01	0.01

Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%