# From wires to partners: how the Internet has fostered R&D collaborations among 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
CP145 - 21, avenue F.D. Roosevelt
1050 Brussels
Belgium
Nicolas.van.Zeebroeck@ulb.ac.be

**February 2010**

*PRELIMINARY AND INCOMPLETE: PLEASE DO NOT QUOTE OR CITE WITHOUT PERMISSION*

## Abstract

This paper studies how IT investments shape the geography of firm innovation. We focus on the role of investments by US firms in basic internet technology (lowering communication costs) on the organization of innovation. We combine this establishment-level IT investment data with data on US patenting activity at the MSA level. Our difference-in-difference econometric estimation approach compares the citation-weighted count of co-invented patents between two firm locations before basic Internet technology diffused (i.e., 1992) to their count of patents after its diffusion (i.e., 1998). For comparison, we also study the effects of Internet investment on patenting behavior within a single firm location. Our results show that when two establishments adopt Internet technology, the number of collaborations between them increases compared to an otherwise identical pair without Internet technology. In contrast, we find that adoption of Internet technology has no impact on the number of research collaborations within a firm location. We find that both results remain robust to numerous specifications and changes to controls.

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

[*] We thank Avi Goldfarb and participants at the ZEW-KMRC Workshop on Social Software for helpful comments and suggestions. 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

How has the diffusion of communication technologies like the Internet influenced research collaborations and the production of new knowledge within firms? This question speaks to two central problems in the economics of organization and economics of technical change. First, it speaks to a large literature on the organization of innovation within organizations (e.g., Azoulay and Lerner, Forthcoming; Cohen and Levin 1989) that has studied how innovation can be organized most effectively. Second, it advances a large literature that has examined how improvements in the availability of information and the lowering of communication costs has shaped the organization of economic activity within firms (e.g., Brynjolfsson and Hitt 2000).

One question of particular interest is how improvements in information technology (IT) have the potential to reshape the geography of firm research collaboration and innovation within firms. Perhaps surprisingly, this issue has thus far received relatively little systematic empirical study. Recently, there has been increasing interest in understanding how IT may shift the geography of academic research collaborations (e.g., Agrawal and Goldfarb 2008; Winkler, Levin, and Stephan Forthcoming; Rosenblat and Mobius 2004; Cummings and Kiesler 2007), however as yet there has been little work examining how IT investments may influence the geography of firm research. This gap in knowledge is significant, given the longstanding interest in the geography of industrial innovation from a research and policy perspective (e.g., Jaffe, Trajtenberg, and Henderson 1993).

In this paper we take a first step toward studying how IT investments shape the geography of firm innovation. To do this, we focus on the role of firm investments by US firms in basic internet technology. Basic internet technology involves the adoption of basic communications such as email use, Internet browsing, and document sharing (Forman, Goldfarb, and Greenstein 2005). In focusing on a specific set of technologies, we are able to isolate the impact of lower communication costs on the organization of innovation. We combine this firm-level IT investment data with data on US patenting activity from the US Patent and Trademark Office (USPTO).

Our econometric approach compares the number of co-invented patents between two firm locations before basic Internet technology diffused (i.e., 1992) to the number of patents after its diffusion (i.e., 1998). That is, we use a difference-in-difference econometric estimation approach to identify the relationship between Internet investments and the pattern of research collaborations. For comparison, we also study the effects of Internet investment on patenting behavior within a single firm location. Our sample period addresses a time period over which Internet technology had diffused but before enough time had elapsed for firms to change the internal organization (in particular, the geographic locations) of its research organization.

Our first set of results assumes that Internet adoption is exogenous to research collaborations. Our results show that when two establishments adopt Internet technology, the number of collaborations between them increases significantly compared to an otherwise identical pair without Internet technology. In contrast, we find that adoption of Internet technology has no impact on the number of research collaborations within a firm location. We find that both results remain robust to numerous specifications and changes to controls.

We next address the assumption that Internet adoption is exogenous. In particular, we address the most likely source of concern: omitted variable bias at the establishment level. We first utilize the timing of Internet adoption as the source of a falsification exercise. We find no evidence that cross-location research collaborations (1990-1994) prior to the diffusion of the commercial Internet were correlated with establishment's later adoption of Internet technology (i.e., in 1998). We next show that our results are robust to the use of instrumental variables estimates that use cost shifters of Internet technology as instruments.

Our findings contribute to several fields of research. First, as noted above, our results contribute to a nascent set of results on the effect of IT use on research collaborations among academic researchers. In particular, the paper most closely related to ours is Agrawal and Goldfarb (2008), who show that adoption of an earlier communication technology, Bitnet, facilitated cross-institution collaboration. However, adoption of Bitnet facilitated an increase in research productivity in particular among middletier institutions that were co-located with large research universities. In contrast, we find that adoption of basic Internet technology 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 argue that the geographic pattern of our results is due to differences in the way that firm and academic research collaborations are formed.

Our research is also related to work on how the use of information technology influences the organization of firms. Theoretical work on this subject has a long history, however researchers have only relatively recently begun systematically testing many theories (e.g., Hubbard 2000; Bresnahan, Brynjolfsson, and Hitt 2002; Bloom et al 2009). Empirical work in this area has most frequently studied how lower information processing and communication costs associated with IT use has influenced the location of decision rights within and between firms. In contrast, our work shows how IT use influences the geography of research collaborations. As we show how IT use facilitates the development of new research teams within an organization, our research most directly informs prior theory research on team-building in this literature (e.g., Marschak and Radner 1972).

Last, 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). Increasingly researchers have presented evidence on the extent to which research activity has globalized (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 this evidence.

### 1.1. Related Research

*Collaborations and productivity*

An abundant literature has evolved that has focused on the association between R&D collaborations of various forms and different knowledge processes or stocks. From the pioneering works by Mansfield and Teece, it appeared that knowledge does not flow easily, and is actually costly to diffuse, absorb and recombine. Kogut and Zander (1992) in particular, showed that knowledge does not easily cross firm boundaries and is easier to diffuse internally. Singh (2005) observed that intraregional and intra-firm knowledge flows are stronger than those across regional or firm boundaries. Cohen and Levintahl (1990) explain this difficulty in transferring knowledge with the notion of absorptive capacity, which is understood as a firm's ability to recognize, value, and assimilate new external information. They suggest that this capacity is largely a function of the firm's level of prior related knowledge, and that cognitive sources of individual absorptive capacity include related knowledge and diversity of background. Cockburn and Henderson (1998) examine the interface between for-profit and publicly funded research in pharmaceutical industry, and observe that firms must exhibit substantial absorptive capacity to capture and appropriate rents to publicly available knowledge.

### *Computer-mediated communications and collaborations*

Why internet and other computer-mediated communication are important is because knowledge is not only difficult to transfer across institutions, but also across geographic distance. Several authors found indeed strong evidence of the localization of knowledge spillovers (e.g. Jaffe et al., 1993; Trajtenberg and Jaffe, 1996; Almeida and Kogut, 1999; Agrawal et al., 2003; Link and Scott, 2003; Thompson, 2006), and one of the key benefits from computer networks is precisely to reduce the costs of coordination and the need for physical co-location between co-workers (Finholt and Sproull, 1990). By observing the Internet adoption patterns of geographically concentrated and dispersed firms, Forman (2005) suggests that the Internet helped firms to reduce communication and coordination costs created by geographic distance.

An extensive literature has therefore focused on the impact of computer networks on scientific collaborations between academic institutions and observed that computer-mediated communication is associated with an increase in the number, size, structure, geographic dispersion and productivity of collaborations (e.g. Kerr and Hiltz, 1982; Finholt and Sproull, 1990; Sproull & Kiesler, 1991; Hesse et al., 1993; Bishop, 1994; Rice, 1994; Kaminer and Braunstein, 1998). In particular, Walsh and Bayma (1996) showed that the increase in scientific (international) collaborations has been facilitated by Internet-related technologies, Walsh et al. (2000) found that the use of computer-mediated communication is positively associated with the productivity and collaboration of academic scientists, and Walsh and Maloney (2002) observed that Internet technologies enabled a change in structure and productivity of academic collaborations. However, to our knowledge, no work has examined the effects of IT investments on collaborations within firms as we do.

## 2. Internet Investment and the Geography of Knowledge Production

Our interest in this paper is examining how a change in communication costs may facilitate a change in the geography of knowledge production. We examine a margin of IT investment that will lower the costs of communication while requiring relatively little in the way in the upfront organizational change. We label this margin basic Internet investment. Basic Internet investment includes such applications as e-mail, web browsing, and passive document sharing.[1] Further, we study a time period prior to when firms have the ability to adjust the location of researchers in response to these lower communication costs. In that sense, our setting allows us to examine directly the effects of lower communication costs on collaboration patterns, holding the location of workers fixed.

It is well documented that the diffusion of electronic communication technologies may have an ambiguous impact on the geography of economic activity (Gaspar and Glaeser 1998; Rosenblat and Mobius 2004; Agarwal and Goldfarb 2008; Van Astyne and Brynjolfsson 2007). In particular, electronic and face-to-face communications may be complements rather than substitutes if collaborations and friendships are usually started through face-to-face communication (Gaspar and Glaeser 1998; Charlot and Duranton 2006). Similarly, electronic communication may reinforce existing social networks or

[1] This margin of investment has been labeled *participation* in prior work (Forman, Goldfarb, and Greenstein 2005). Basic Internet differs from other margins of investment that researchers have explored to study the benefits of IT that involve up-front changes to business processes. Such margins of investment have sometimes been labeled process-enabling IT (Brynjolfsson, McAfee, Sorrell, and Zhu 2008). Our margin shares commonalities with the set of applications that have been labeled as network IT in prior research (McAfee 2006)—that may facilitate changes to firm organization but does not require such changes up front.

communities with similar interests (Rosenblat and Mobius 2004). These mechanisms would tend to reinforce communication patterns along existing geographic lines.

However, a key feature of many models on the collaborative effects of IT is that the decision of whom to collaborate or communicate with is mediated by existing geography or interests, and that electronic communications strengthen these established relationships. In contrast, we examine a setting where ex ante relationships –and the decision of who to collaborate with—are less likely to be determined autonomously by individual agents based on geography or interests.

In particular, our first set of analyses examines the effect of basic Internet investment on the geography of collaborations within firms. In this case, the composition of research teams will be determined to maximize knowledge production, accounting for on the one hand the advantages of things like economies of scale, scope, and spillovers (e.g., Henderson and Cockburn 1996)—which would encourage the formation of larger teams with deeper and more diverse competencies—and the communication and coordination costs of large, geographically dispersed teams—which would tend to limit the size and dispersion of such teams (Marschak and Radner 1972).

By reducing the costs of communication, the introduction of electronic communication technologies like basic Internet reduces the coordination and communication costs of large, geographically dispersed teams. By removing an important cost to team formation, such technologies may influence team formation in a number of ways in the short run. Here we focus on one particular implication: the effect on geographically dispersed teams. We argue that while adoption of basic Internet will lower the incremental costs of another team member whether in the same or different location, the costs of adding an additional team member from another location will be particularly affected. That is, while the adoption of basic Internet will increase the size of same location collaborations—it will have its greatest impact on increasing the number and size of cross-location collaborations.

## 3. Data

We use a variety of data sources to describe how adoption of basic Internet influences research collaborations. We describe each of these below. Descriptive statistics are provided in Table 1.

**Patent Data.** To measure the impact of basic Internet adoption of the geography of knowledge production within firm, we match our IT investment data with data on patents filed with the US Patent and Trademark Office (USPTO). To match patent data to our IT data, we require consistent information on the identity of the firm. For this purpose, we use the NBER Patent Data Project's matching data set (Hall et al. 2005) which maps patents to a consistent set of unique firm identifiers based on the "GVKEY" identifier from the COMPUSTAT database. We obtain the universe of patents with a matching GVKEY that were applied for during 1990-2000.

As is well known, US patent data contain information on inventor locations but not on the locations of the firms where research is performed. To obtain a consistent measure of both IT inputs and patent outputs, we aggregate both our patent and IT data to the level of Metropolitan Statistical Areas (MSAs). In cases where Consolidated Metropolitan Statistical Areas (CMSAs) were present, we used those.[2] 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. MSAs are constructed on the basis of commuting patterns and are widely used as a unit of geographic economic activity. In short, we group patents and IT into firm-MSA-years.

Our procedure for mapping patents to firm-MSA-years first identifies the patent firm based on the NBER Patent Data project and MSA based upon the zip code of the inventor (obtained through the USPTO Patents BIB data product). 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 (2005). In particular, we consider only citations within five years of the grant to avoid truncation bias, and deflate the citations received by each patent by its IPC4-year average.

[2] 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) and while the results are qualitatively similar, they are somewhat weaker due presumably to commuting patterns of inventors across PMSAs within the same CMSA.

**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 new forms of IT (Bresnahan and Greenstein 1996; Forman 2005; Forman, Goldfarb, and Greenstein 2005) and the productivity implications of IT investment (Brynjolfsson and Hitt 2003; Bresnahan, Brynjolfsson, and Hitt 2002; Bloom, Sadun, and Van Reenen 2007; Bloom, Garicano, Sadun, and Van Reenen 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 advanced Internet applications, 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 to predominately large establishments, our algorithm for matching our IT data to the patent data which draws upon the NBER Patent Data Project similarly requires us to focus upon large firms. Thus, our analysis should be viewed as a study of IT and research collaborations within large firms. The primary firms in this data set are well established and have existed well prior to the diffusion of the commercial Internet. That is, our data are not in general small firms whose emergence coincided with the diffusion of the Internet.

We focus on those facets of Internet technology that became available only after 1995 in a variety of different uses and applications. 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 basic communication technologies that involve little adaptation by users to be implemented successfully. In particular, we define an establishment as a basic Internet adopter if it indicates that it has one of the following: basic access, an intranet, or uses the internet for research purposes.

To map our establishment-level IT data to unique firm-MSA-years, we map the unique firm identifier in the Harte Hanks database to the GVKEY provided in the NBER Patent Data Project. We then assign establishments to MSAs using their zip code. In the case where there are multiple establishments in the same MSA, we use the first incidence of adoption by one of the establishments in the MSA to denote internet adoption. For our analysis is data set, we include only firm-MSA-year triplets that are from manufacturing firms (SIC 2000-4000) and are in triplets with at least one patent over the period 1992-1998.

**Firm-MSA pairs**. The focus of our study is on the effects of IT investment on within and cross-location research collaboration. Our primary analysis studies how adoption by both Firm-MSA locations in a pair influences the number of collaborations between the two locations, and whether adoption within a particular pair influences the number of collaborations within a location. Given that our set of analyses are within firm, we form the complete set of pairwise combinations of establishments 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, and define a pair as having adopted Internet if both locations are adopters.

**Other controls**. We combine these data with level information from a number of sources. 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-establishment R&D flow dollars by normalizing total spending by the number of establishments in our data.

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. These measures are computed at the county level and then matched to MSAs. We then compute the average of these across MSAs in the collaboration pair. First, based on County Business Patterns data, we compute the percent of manufacturing employment in the MSA, the average weekly wage in the MSA, and the log of MSA employment. Second, using the USPTO data, we compute the total number of patents in the MSA-year.

## 4. Empirical Strategy

To measure the impact of Internet on collaborations between firm-location pairs, we use a difference-in-difference identification strategy, comparing the number of (citation-weighted) collaborations of a time period before basic Internet technology diffused (1992) to those of a period where we observe adoption (1998). Our endogenous variable will be $Patents_{it}$, which represents the number of collaborations as measured by co-invented patents for a particular firm pair $i$ in patent application year $t$.[3] Internet technology had not diffused among firms priors to 1995 except in very rare cases, so we set the value of this variable to zero in 1992. This yields the following regression equation:

$$Patents_{it} = \alpha_1 X_{it} + \alpha_2 Z_{it} + \beta Internet_{it} + \mu_i + \tau_t + \varepsilon_{it}$$

Here, $\mu_i$ is a firm-pair fixed effect that gets differenced out in the estimation, and $\tau_t$ is a time dummy that captures changes in average levels of firm-pair collaboration over time.[4] The variable $Internet_{it}$ measures whether both establishments in the pair adopted basic Internet. We have two types of controls: the variables in $X_{it}$ capture firm-pair controls for things like R&D expenditures and establishment size that may affect the volume of collaborations in a firm-pair. The variables in $Z_{it}$ capture location level characteristics of the pair that may influence innovation. We have assumed that $\varepsilon_{it}$ is a normal i.i.d. variable.[5]

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 collaborations, as proxied by co-invented patents: a test of β>0 against the null of β=0.

[3] The median number of collaborations is zero and the mean is 0.26, so the level of citation-weighted patents rather than the log was judged to be a more appropriate measure. However, we have also estimated the model using the log of patents and the results are qualitatively similar.

[4] Since we treat our standard errors appropriately, note that this is exactly equivalent to a two-period difference estimator.

[5] Since we estimate the standard errors using heteroskedasticity-robust methods, the two-period framework is especially appealing. Stock and Watson (2008) show that the standard fixed-effects heteroskedasticity-robust variance matrix estimator is inconsistent if T is fixed and greater than 2.

To measure the impact of basic Internet adoption of within-location collaborations, we estimate a variant of the above equation for establishments j (collocated within the same MSA),

$$Patents_{jt} = \alpha_1 X_{jt} + \alpha_2 Z_{jt} + \beta Internet_{jt} + \mu_j + \tau_t + \varepsilon_{it}$$

Here, $Patents_{jt}$ represents the number of co-invented patents in location $j$, $Internet_{jt}$ is a binary of whether basic Internet has been adopted at the location, and $X_{jt}$ and $Z_{jt}$ are firm-location and location level controls, respectively. Here the hypothesis is that adoption of basic Internet within the firm-location will be associated with an increase in the number of within-location collaborations: again, a test of β>0 against the null of β=0.

In both regressions, we assume that the unobservable determinants of patents can be decomposed into an additively separable fixed component and a time-varying component that is constant across firm-pairs (or firm-locations). To start, we also assume that there are no unobserved factors in $\varepsilon_{it}$ that are correlated with Internet adoption.

We then explore this latter assumption: a particular concern is that unobserved features of the firm establishments in the pair or their locations may be correlated both with Internet adoption and patent growth. In particular, we do two things to explore this assumption. First, we present instrumental variable estimates that use measures of local telecommunications costs, adoption by competitors, and programming capabilities in related locations as instruments for Internet investment. Second, we perform a falsification test of whether cross-location research collaborations (1990-1994) prior to the diffusion of the commercial Internet were correlated with establishment's later adoption of Internet technology (i.e., in 1998).

## 5. Results

We first establish a relationship between Internet adoption and the number of cross-MSA research collaborations measured through patents. We demonstrate that this result is robust to a variety of specifications, and to the use of instrumental variables. We next demonstrate that Internet adoption is not associated with increased collaborations within an MSA.

In Table 2, we show the baseline results across cross-location pairs. Column 1 shows the correlation between Internet adoption and collaborations without any pair or time fixed effects, and without any time dummies. There is a strong correlation in these results between Internet adoption and cross-location patenting. In Column 2 we include our baseline results, which includes both pair and time fixed effects along with our complete set of controls. While the size of the coefficient estimate drops almost by half, the results are still statistically (at the 5% level) and economically significant. If both establishments in the pair have basic Internet this translates into a 0.11 increase in the number of weighted patents. When compared to the mean number of 0.26 patents per pair-year, this translates roughly into an increase of 42.3%. We explore further robustness in columns (3) and (4). Column (3) shows that our results continue to hold when we use only MSAs and exclude our "phantom" MSAs that are outside of metropolitan areas. Column (4) shows that our results continue to hold when we use the log of collaborations.

In column 5 we show the results of a falsification test to explore the extent to which our results may be influenced by omitted variable bias. This falsification test utilizes the timing of Internet adoption. As has been reported extensively elsewhere, the commercial Internet diffused rapidly beginning in the end of 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 to our results and circumscribes the way in which omitted variable bias may be influencing our results.

Column (5) shows that there is little impact on Internet adoption over the period 1990-1994: the coefficient on Internet adoption is small (0.0297) and insignificantly different from zero.

To further examine the robustness of our results in Table 3 we include the results of instrumental variable estimates. We have four instruments in total for one endogenous variable. We include two variables to proxy for local deployment costs: the year in which the local state capped prices that incumbent local exchange carriers (ILECs) could charge entrants and the year in which they switched to rate of return regulation.[6] By influencing the local costs of deployment, these variables should be correlated with local Internet adoption. However, it is very unlikely they will be correlated with growth in patenting. Finally, we use the multi-establishment nature of the firms in our data to construct two further sets of instruments. We measure the total number of programmers in other establishments and other counties, but in the same firm. We use the average as an instrument. Forman, Goldfarb, and Greenstein (2008) show these variables are correlated with Internet adoption. They are also likely to uncorrelated with growth in patenting; our programmers variable reflects the presence of IT skills in linked counties. Last, we use the average adoption rate of competing firms in other locations in which the firm has establishments. Because firms benchmark their IT investments with competitors (e.g., Cortada 1997), these adoption rates are likely to be correlated with Internet adoption. However, because they represent adoption of other firms in linked counties, they are very unlikely to be correlated with patenting activity in the firm-pair.

We follow the strategy suggested by Angrist and Pischke (2009) in presenting our results and addressing concerns about weak instrument bias. Column (1) shows our first stage results and demonstrates that our instruments are statistically significant separately, and an F-test shows that they are also significant jointly (F-test 23.44, p-value 0.0000). Column 2 shows our baseline instrument results and show that our results are robust to the use of these instruments; if anything, they are stronger. An overidentification test on these estimates does not reject the null hypothesis that our instruments are orthogonal to the 2nd stage residuals ($x^2$=6.1793, p-value 0.1032). We also present estimates using our best instrument, which will be median unbiased. Here the first-stage results show the instrument is again

---

[6] We thank Avi Goldfarb and Shane Greenstein for providing these instruments to us.

statistically significant, and the second-stage results remain significant at the 5% level. While we use 2SLS estimates here, our results are also robust to LIML estimation.

In Table 4 we show the results of our model that explores the correlation between Internet adoption and within location adoption. While column (1) suggests that without fixed effects there may be some correlation between Internet adoption and patenting, this relationship disappears once we add location and time fixed effects. We do not find a relationship between Internet adoption and patenting in any of our models.

## 6. Conclusion

This study focuses on the effect of basic internet adoption on the shape and organization of R&D collaborations. To do so, we match IT-investment data with patenting activity at the USPTO, both aggregated at the firm-MSA-year level. With a difference-in-difference econometric estimation approach, we find robust empirical evidence that Internet adoption has fostered the citation-weighted number of co-invented patents between pairs of locations within a firm. On the contrary, we find no evidence of such a link on within-location patenting. These results suggest therefore that Internet adoption has fostered collaborative R&D projects within firms and has therefore led to a wider geographical dispersion of innovation activities.

These findings have some important implications in terms of R&D organization and innovation performance, as they suggest that lower communication costs in the private sector may lead to a jump in distant collaborations more than in local innovative outputs. They nonetheless open the door for further research to better understand how and where Internet investments most affect collaborative R&D. First, one may wonder whether internet investments have lead to any change in the size of local R&D teams. The absence of a jump in within-MSA output overall does indeed not preclude a significant effect of Internet on internal R&D organization and local collaborations. Second, we also need to understand better whether our results show any significant difference in magnitude across different dimensions such as industry or location (e.g. population density, size, innovation productivity, distance, etc.). Another question to investigate is whether Internet investments impact collaborations differently across similar v. dissimilar fields of research, or when research activities are more basic v. applied. Finally, our results do not yet allow us to determine whether the marginal gains in co-invented patent output are mostly coming from (1) whether firm establishments that didn't collaborate previously now collaborate after adoption of the Internet; (2) whether places that did collaborate now collaborate more; or (3) whether the value (forward citations) of co-invented patents has increased.

**References**


Azoulay, P. and J. Lerner (Forthcoming) Technological Innovation and Organizations, in the Handbook of Organizational Economics, eds. Gibbons, R. and J. Roberts, Princeton University Press.

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

Almeida, P. and B. Kogut (1999), Localization of knowledge and the mobility of engineers in regional networks, Management Science, 45(7), pp. 905-917.

Angrist, J. D. and J.-S. Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton, NJ: Princeton University Press.

Bishop, A. (1994), The role of computer networks in aerospace engineering, Library Trends, 42, pp. 694-729.

Bloom, N., R. Sadun, and J. Van Reenen. 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.

Bresnahan, T., E. Brynjolfsson, and L. Hitt (2002), Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence, 117(1), 339-376.

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.

Brynjolfsson, E., A. McAfee, M. Sorell, and F. Zhu (2008), Scale Without Mass: Business Process Replication and Industry Dynamics, Harvard Business School Technology & Operations Mgt. Unit Research Paper No. 07-016

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, The Journal of Industrial Economics, 46(2), pp. 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, chapter 18.

Cohen, W. and D. Levinthal (1990), Absorptive Capacity: A New Perspective on Learning and Innovation, Administrative Science Quarterly, 35(1), pp. 128-152.

Cortada, J. W. (1997), Best Practices in Information Technology: How Corporations Get the Most Value from Exploiting their Digital Investments, Prentice Hall.

Cummings, J. and S. Kiesler (2007), Coordination costs and project outcomes in multi-university collaborations. Research Policy 36, 1620-1634.

Finholt, T. and L. Sproull (1990), Electronic Groups at Work, Organization Science, 1(1), pp. 41-64.

Forman, C. (2005), The Corporate Digital Divide, Determinants of Internet Adoption, Management Science, 51(4), pp. 641-654.

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

Forman, C., A. Goldfarb, and S. Greenstein. 2008. Understanding the Inputs into Innovation: Do Cities Substitute for Internal Firm Resources? *Journal of Economics and Management Strategy* 17(2): 295–316.

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

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

Hall, B., A. Jaffe, M. Trajtenberg (2005), Market value and patent citations, The RAND Journal of Economics, 36(1), 16-38.

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.

Hesse, B., L. Sproull, S. Kiesler and J. Walsh (1993), Returns to science, Communications of the ACM, 36, pp. 90-101.

Hubbard, T. (2000), The Demand for Monitoring Technologies: The Case of Trucking, The Quarterly Journal of Economics, 115(2), 533-560.

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), pp. 577-598.

Jorgenson, D. and K. Stiroh (1995), Computers and Growth, Economics of Innovation and New Technology, 3(3&4), pp. 295-316.

Jorgenson, D. and K. Stiroh (1999), Information Technology and Growth, American Economic Review, 89(2), pp. 109-115.

Kaminer, N., and Y. Braunstein (1998), Bibliometric analysis of the impact of Internet use on scholarly productivity, Journal of the American Society for Information Science, 49, pp. 720-730.

Kerr, E. and S. Hiltz (1982), Computer-mediated communication systems, Academic, New York.

Kleis, L., P. Chwelos, R. Ramirez, and I. Cockburn (2009), 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), pp. 383-397.

Link, A. and J. Scott (2003), U.S. science parks: the diffusion of an innovation and its effects on the academic missions of universities, International Journal of Industrial Organization, 21, pp. 1323-1356.

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

Marschak, J. and R. Radner (1972), Economic Theory of Teams, New Haven: Yale University Press.

McAfee, A. (2006), Mastering the Three Worlds of Information Technology, Harvard Business Review, November 1, 2006.

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

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

Sproull, L. and S. Kiesler (1991), Connections, MIT Press, Cambridge, MA.

Stock, James H. and Mark W. Watson (2008), Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression, Econometrica, 76(1), 155–74.

Thompson, P. (2006), Patent citations and the geography of knowledge spillovers, Review of Economics and Statistics, 88(2), pp. 383-388.

Trajtenberg, M. and A. Jaffe (1996), Flows of Knowledge from Universities and Federal Labs: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundaries, NBER Working Paper No. W5712.

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

Walsh, J. and N. Maloney (2002), Computer Network Use, Collaboration Structures, and Productivity, in Distributed work, Hinds, P. and S. Kiesler (eds.), MIT Press, MA, p. 433.

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

Table 1: Descriptive statistics for dependent variables, IT measures, and instruments (for 1998)

| Variable | Mean | Std. Dev. | Minimum | Maximum | Number of obs |
|---|---|---|---|---|---|
| Number of (weighted) collaborations | 0.2566 | 2.2021 | 0 | 106.66 | 5878 |
| Has Internet in both locations | 0.6899 | 0.4626 | 0 | 1 | 5878 |
| Log of per-establishment R&D expenditures | 3.1763 | 1.5149 | -0.9715 | 7.7295 | 5878 |
| Percent manufacturing employment in MSA | 0.1973 | 0.0644 | 0.0391 | 0.4861 | 5878 |
| Average weekly wage in MSA | 605.00 | 87.3969 | 382.6797 | 848.329 | 5878 |
| Log of MSA employment | 13.8355 | 0.9511 | 10.3316 | 15.7005 | 5878 |
| Number of patents in MSA | 1682.699 | 1745.817 | 1.5 | 9240 | 5878 |

Table 2: Multi-Location Collaborations Increase with Internet use

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | No fixed effects | Baseline | MSAs only | Log of collaborations | 1990-1994 |
| Internet | 0.1756** | 0.1057 | 0.1284 | 0.0154 | 0.0297 |
| | (0.0536) | (0.0459)* | (0.0683)+ | (0.0085)+ | (0.0290) |
| | | | | | |
| Observations | 16751 | 14177 | 9557 | 14177 | 10976 |
| (within) $R^2$ | 0.0082 | 0.0109 | 0.0129 | 0.0126 | 0.0064 |
| Fixed Effects | No | MSA-Pair Time | MSA-Pair Time | MSA-Pair Time | MSA-Pair Time |
| Controls | SIC dummies | All | All | All | All |

Dependent variable is citation-weighted number of collaborations (in column 5 log of collaborations). Unless otherwise stated, years are 1992 and 1998. Controls include year dummy, log of establishment R&D spending in pair, log of total employment in pair, percent manufacturing employment in pair, average weekly wage in pair, log of average employment in pair, and number of patents in pair. Robust standard errors in parentheses.

Table 3: Results are Robust to Use of Instrumental Variables

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | First stage results for baseline IV | Baseline $2^{nd}$ Stage IV results (2SLS) | First stage results for single instrument IV | $2^{nd}$ Stage IV results for single instrument (2SLS) |
| Internet | | 1.0312 | | 1.9158 |
| | | (0.4076)* | | (0.7832)* |
| Log of establishment | 0.0419** | 0.0343 | 0.0405 | -0.0025 |
| R&D | (0.0116) | (0.0544) | (0.0116)** | (0.0659) |
| Log of total pair employment | 0.0048 | -0.0611 | 0.0068 | -0.0738 |
| | (0.0155) | (0.0807) | (0.0154) | (0.0847) |
| Percent Manufacturing in | -0.1719 | 1.2965 | 0.7352 | 1.1058 |
| MSA | (0.5556) | (1.5549) | (0.5384) | (1.7477) |
| Average weekly wage in MSA | -0.0001 | 0.0019 | 0.0001 | 0.0016 |
| | (0.0003) | (0.0014) | (0.0003) | (0.0015) |
| Log of total employment in | -0.0225 | -0.2360 | -0.1277 | -0.2664 |
| MSA | (0.1172) | (0.4942) | (0.1164) | (0.5272) |
| Number of patents in MSA | 0.0000 | 0.0001 | 0.0000 | 0.0001 |
| | (0.0000) | (0.0000)+ | (0.0000) | (0.0000)* |
| Programmers in other | -0.0001 | | | |
| Locations | (0.0000)** | | | |
| Competition in Other | 0.0154 | | | |
| Locations | (0.1129)** | | | |
| First change to price cap | 0.0188 | | 0.0193 | |
| Or freeze | (0.0035)** | | (0.0036)** | |
| First change to ROR | -0.0170 | | | |
| Regulation | (0.0025)** | | | |
| Constant | 32.6984 | -0.8595 | -1.1311 | -1.4216 |
| | (5.0487) | (0.3007)** | (0.0036)** | (0.5138)** |
| Number of observations | 5878 | 5878 | 5878 | 5878 |

Instruments are number of programmers in other locations where the firm has establishments, adoption of internet by other firms in the same industry in other locations where the firm has establishment, a dummy indicating whether the state is the first to change to a price cap or freeze, and another dummy that indicates the state is the first to change to rate of return regulation. All regressions are run between 1992 and 1998. Robust standard errors in parentheses.

Table 4: Little evidence that single MSA collaborations increase

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | No fixed effects | Baseline | MSAs only | Log of collaborations | 1990-1994 |
| Internet | 5.2397** | -1.6937 | -2.3598 | -0.0277 | 0.9967 |
| | (2.2960) | (3.2959) | (3.9528) | (0 .0766) | (1.1480) |
| | | | | | |
| Observations | 3236 | 2609 | 2164 | 2609 | 2384 |
| (within) $R^2$ | 0.0189 | 0.0663 | 0.0710 | 0.0629 | 0.0322 |
| Fixed Effects | No | MSA-Pair Time | MSA-Pair Time | MSA-Pair Time | MSA-Pair Time |
| Controls | SIC dummies | All | All | All | All |

Dependent variable is number of weighted collaborations (in column 5 log of collaborations). Unless otherwise stated, years are 1992 and 1998. Controls include year dummy, log of establishment R&D spending in pair, log of total employment in pair, percent manufacturing employment in pair, average weekly wage in pair, log of average employment in pair, and number of patents in pair. Robust standard errors in parentheses.