Information technology and the distribution of innovative activity

By Chris Forman, Avi Goldfarb, and Shane Greenstein

Extended Abstract

Vannevar Bush's publication *Science: The Endless Frontier* lays out a range of institutions and mechanisms to promote innovation. Among its many major themes, this key policy work distinguishes between the source information that acts as input for innovation, the production of innovation, and the process of dissemination. This framework generally stresses the importance of the communication and dissemination of information for the production of innovation, and the importance of dissemination for the use of innovation, post-invention.

That conceptualization and framework has long been a staple of analytical work. It frames a range of questions about the localization of information, and related questions about how the costs of knowledge transmission, dissemination, and collaboration increase in distance (Jaffe, Trajtenberg, and Henderson 1993; Glaeser, Kerr, and Ponzetto 2010; Delgado, Porter, and Stern 2008; Saxenian 1994, etc.). That conceptualization also frames a range of questions about the collocation of innovative activity. Accordingly, the creation and maintenance of geographic clusters of innovation, and its links to regional economic growth, have been an important part of innovation policy.

Increasing use of information and communication technology (ICT) may change the importance of collocation in generating new innovation. It is possible that ICT reduces the geographic concentration of innovation because it allows researchers to communicate across long distances at little cost. It also reduces the requirement that researchers collocate with complementary resources, such as information repositories (e.g., libraries). Indeed, such potential gains partially generated NSF support for building the Internet inside US research universities (Abbate, 1999). When the commercial internet initially began to diffuse in the early 1990s, the anticipated decline in communications costs generated a flurry of optimistic predictions about how its use would change the location of economic activity more broadly. The bestsellers *The Death of Distance* and *The World Is Flat* were just two of the books that espoused that rosy view.

With the benefit of hindsight and calmer consideration, alternative views have developed. Most pointedly, it is also possible that ICT could increase the geographic concentration of innovation. Declining communications are not merely a substitute for human activity, as if the *primary* impact of declining communications costs is to reduce the frequency of face-of-face meetings. The marginal impact of ICT can arise due to substitution with face-to-face communication, but it is possible that the opposite will arise; that it will make face-to-face meetings more productive. ICT is also complementary with many other existing economic activities. It complements existing (physical or human) capital in a variety of activities, and its marginal impact could be higher for the most intense users, namely, skilled labor -- which, for a variety of reasons tends to locate in major cities. It is, therefore, also possible for declining communication costs to enhance activity within cities more than low density locations. In that case, ICT skills may increase the geographic clustering of innovation.

Increasing evidence suggests complements were the norm during the initial diffusion of the internet. For example, the first generation of the commercial internet enhanced the returns to economic activity within dense urban settings comparatively more than in less dense settings. The data (so far) are consistent with complementarities in production between the internet and existing capital, skill biased

technical change, and/or Marshallian externalities. That norm does hide some differences across particular settings, and prior research has shown that the impact of changes in ICT on the geographic distribution of economic activity is ambiguous (Forman, Goldfarb, and Greenstein 2005, 2008, 2012; Agrawal and Goldfarb 2008; Forman and van Zeebroeck 2012; Forman, Ghose, and Goldfarb 2008).

To our knowledge, despite this prior work, there has been no systematic empirical exploration of how the deployment of ICT and its increasing use has changed the geographic distribution of innovation. We consider it an empirical question whether, on balance, the deployment commercial internet, for example, increased or decreased the marginal benefit of innovation in isolated areas relative to areas with high density.

Our study will explore empirically how changes in ICT affect the geography of knowledge production. Combining detailed information on the location of patents by industry with the location of ICT adoption by industry, we will assess how ICT explains changing patterns in the geographic concentration of innovation across industries.

Sketch of methodology

The unit of observation for our core analysis will be the location-industry-year. We will correlate changes in innovative activity in a particular location-industry with changes in ICT in that location-industry, controlling for a variety of location and industry-specific factors.

We will measure ICT using data from Harte Hanks Market Intelligence (as in our prior work, and also as in Bloom, Garicano, Sadun, and Van Reenen 2009 and McElheran 2011). This database contains rich establishment-level data including use of internet applications, industry, and location. We plan to focus on the sample from December 2000, though we will also explore the 1996 and 2006 surveys. Consistent with our prior work, we will focus on two margins of ICT adoption, which we label *basic* and *advanced*. We will also add data on consumer adoption of internet technology using the Computer and Internet Use Supplements to the Current Population Survey from 1999, 2000, 2001, and 2003. These data sets can identify adoption rates by MSA, industry, and occupation.

For our dependent variable, we will focus on three core measures of innovative activity: the total number of granted patents (in a county in an industry—mapped from patent class), the total number of patents granted with more than one inventor, and the total number of "long-distance" patents granted (where a long-distance patent is defined as one where the maximum distance between the inventors is at least 50 miles). The total number of patents will give us a sense of the degree to which ICT might affect the geographic distribution of knowledge production. The total number of long distance patents will allow us to look for a mechanism behind why ICT may affect the geography of knowledge production. For the patent data, we plan to use the NBER patent database.

We will also include a number of controls to address possible spurious correlation between ICT adoption and innovation. In particular, we will control for county-level population and firm demographics using information from the US Census.

First we will focus on growth in innovation as a function of internet adoption. Specifically,

$$(1)\ Log(Y_{ic05}) - Log(Y_{ic95}) = \alpha X_{ic} + \beta Internet_{ic} + \varepsilon_{ic},$$

where Y_{ict} is our measure of innovation (number of patents, collaborative patents, and long distance patents granted) in industry *i* county *c* and year *t*, *internet* is a measure of internet use in county c in

industry *I* (as in our previous work, we exploit the timing of the diffusion of the commercial internet and set internet equal to zero in the base period), and X_{ic} is a vector of controls for population and local business characteristics at the beginning of the sample period (1995 in this case). This will establish the degree to which ICT is associated with innovation.

Next, we explore whether the association of ICT with innovative activity varies by city size and the existence of an industry cluster. In particular,

$$(2a)\ Log(Y_{ic05}) - Log(Y_{ic95}) = \alpha X_{ic} + \beta Internet_{ic} + \phi(Internet_{ic} x HighPopulation_{c}) + \varepsilon_{ic},$$

$$(2b)\ Log(Y_{ic05}) - Log(Y_{ic95}) = \alpha X_{ic} + \beta Internet_{ic} + \phi(Internet_{ic} x Cluster_{ic}) + \varepsilon_{ic},$$

This analysis will examine whether internet use is correlated with a change in the geographic distribution of knowledge production. We view the results of these regressions as the likely contributions of the chapter. They allow us to examine how the marginal impact of ICT varies across locations.

The results of this analysis will have important implications for the structure of innovation policy. If ICT makes it easier for isolated research to make important contributions to knowledge, then innovation policy initiatives that facilitate such contributions will be more effective. In contrast, if ICT increases the propensity to innovate in areas that are already highly innovative, then policy should continue to focus on building clusters.

Current state of the research project

This research project is still in an early phase. We have all the necessary data, and considerable experience running similar analyses. We have not yet completed any analysis.