

## **Agglomeration of invention in the Bay Area: Not just ICT\***

Chris Forman, Georgia Institute of Technology

Avi Goldfarb, University of Toronto

Shane Greenstein, Harvard University

December 2015

### **Abstract**

We document that the Bay Area rose from 4% of all successful US patent applications in 1976 to 16% in 2008. This is partly driven by the increase in the prevalence of information and communication technology; however, even for patents unrelated to information and communication technology, we see a disproportionate increase in the share of all US patents from the Bay Area. We interpret this to suggest that there has been a trend to coagglomeration in invention across technologies. We explore several possible explanations for this trend, and conclude that the size of firm or simple measurement error cannot explain it.

JEL Code: R12, O31

---

\* Max Koven and Yasin Ozcan provided excellent research assistance. We thank seminar participants at the University of Toronto for helpful comments. Avi Goldfarb thanks SSHRC for research support.

## 1. Introduction

Does invention agglomerate, and if so, where does it agglomerate? In this paper we examine changes in patterns of agglomeration in invention over time, using data on all US patent applications.

There are plenty of reasons to expect invention to agglomerate. Carlino and Kerr's (2015) recent handbook chapter summarizes many such results, emphasizing the role of input sharing, labor market matching, and knowledge spillovers, among others. Knowledge spillovers received an especially large fraction of attention in their chapter, and in the literature overall (e.g., Audretsch and Feldman 1996; Agrawal and Henderson 2002; Moretti 2004, 2012; Rosenthal and Strange 2008; Jaffe, Trajtenberg, and Henderson 1993; Thompson and Fox-Kean 2005; Kerr and Kominers 2014; Moser 2011; Waldinger 2012; Azoulay et al 2010).

Simple economics might forecast that most invention agglomerates in the same area as the primary using industry (Carlino and Kerr 2015). For example, patents related to automotive technology are clustered in Detroit (Hannigan, Cano-Kollmann, and Mudambi 2015). Or causality could be reversed: The location of a break-through invention can lead to industry agglomeration and localized follow-on invention (Duranton 2007; Kerr 2010). We label this “colocation” between invention and industry.

However, other forces push away from colocation. Invention itself is an economic activity and it shares inputs, such as specialized labor institutions, particular intellectual property contracts, and information spillovers from one type of invention to another. If such forces are strong, they could lead to agglomeration of lots of different types of invention in one place. We call this “coagglomeration of invention” For many industries, the key inventions could be in a location distinct from the place where production for the downstream using industries reside.

Using patent data to measure invention, there are two approaches to investigate colocation and coagglomeration of invention. One is to map the agglomeration of downstream industries and invention and measure the geographic correlation. We take another approach. We look for evidence of coagglomeration of invention – namely, invention from distinct areas appearing in the same location, irrespective of downstream using industry.

We find evidence consistent with the hypothesis of coagglomeration. We demonstrate a strong trend toward the clustering of patenting to the San Francisco Bay Area from 4% of US patents in 1976 to 16% of US patents in 2008, a time period when the fraction of the US population in the Bay Area did not increase substantially relative to the US population as a whole.<sup>1</sup> While this increase in Bay Area patenting is partly driven by the increasing fraction of patents in information and communication technologies (ICTs), ICTs cannot fully explain the trend. The San Francisco Bay Area has seen a substantial increase in its share of patents, even for patents that seem quite distant from ICTs. Our broadest definition of ICT patents includes all patents in information technology, communications technology, electronics, a broad measure of software based on patent classes and a textual search of patent titles and abstracts, and all patents that cite any of these patents. The remaining non-ICT patents were 66% of all patenting in 1976 but just

---

<sup>1</sup> According to the US Census, the Bay Area grew from 2.54% of the US population in 1980 (5.74 million residents) to 2.65% of the US population in 2010 (8.15 million residents).

26% in 2008. In 2008, 6.2% of such patents have inventors based in the Bay Area second only to New York City's 8.1%.<sup>2</sup>

Our results are consistent with coagglomeration of invention in the Bay Area. While others have documented a tendency toward agglomeration of patenting by industry, we believe we are the first to document a general tendency toward agglomeration in patenting across industries and patent classes. Further, our study is unique in its documentation of agglomeration in one particular region, the Bay Area.

Coagglomeration has been documented in other settings and other industries. For example, Rosenberg (1963) analyzes how sewing machines, bicycles, and automobiles located in Northern Ohio and southeastern Michigan as they shared the same set of inventions in machine tools, and the growing downstream industries induced additional improvements in those innovations over time. Glaeser (2005) discusses coagglomeration of many industries in New York City, starting in the nineteenth century. Summarizing prior literature, he argues that New York's dominance started with shipping. There is a clear reason why New York could dominate in shipping: New York has a particularly appealing natural harbor beside an inland waterway. Shipping led to risk sharing and insurance, which led to finance. Shipping also led to manufacturing and early book publishing. Population growth, combined with manufacturing and finance led to other services. A number of recent researchers have explored the causes and consequences of such coagglomeration, including Ellison, Glaeser, and Kerr (2010) and Helsley and Strange (2015). Of particular relevance to our study, Delgado, Porter, and Stern (2014) document complementarities between employment and patenting in regions with multi-industry clusters.

At this point, our results do not provide a definitive conclusion on the cause of this broad increase in coagglomeration in invention. A variety of mechanisms are possible including regulations such as non-enforcement of non-compete clauses (Franco and Mitchell 2008; Marx, Singh, and Fleming 2015), agglomeration or expertise in startup financing (e.g., Chen, Gompers, Kovner, and Lerner 2010), shared labor markets across invention types (Almeida and Kogut 1999), and knowledge spillovers across invention types.

One important limitation of our analysis is the use of patents as a measure of invention. Patents measure invention imperfectly, and the ease with which they can be measured means that economists have been perhaps overly focused on patents to measure invention. Some patents are more important than others and many inventions are never patented. Still, patents are a useful measure because they are observable, and comparable across time and categories. This will bias our results if the inventors in the Bay Area have become increasingly likely to patent when they invent.

Of course, we are not the first to document the agglomeration of ICT in the Bay Area. Garcia-Vicente et al (2014) show that such agglomeration took place primarily in the 1980s and 1990s. Our results are consistent with this timing. A variety of authors have explored the reasons behind the agglomeration of the ICT industry in the Bay Area and its dynamics in generating new firms and new ideas (Almeida and

---

<sup>2</sup> One unusual aspect of patenting in the San Francisco Bay Area is that invention is not centered in the city but in Silicon Valley. Therefore, while we refer to other cities by the city names, we refer to the "Bay Area" rather than "San Francisco" to describe the San Francisco Consolidated Metropolitan Statistical Area. We use the 2013 definition, which includes the following 12 counties: Alameda (fips 06001), Contra Costa (fips 06013), San Francisco (fips 06075), San Mateo (fips 06081), Marin (fips 06041), Santa Clara (fips 06085), San Benito (fips 06069), San Joaquin (fips 06077), Sonoma (fips 06097), Solano (fips 05095), Santa Cruz (fips 06087), and Napa (fips 06055).

Kogut 1999, Kerr and Kominers 2015, Saxenian 1994; Franco and Mitchell 2008; Marx, Singh and Fleming 2015, etc.). Our contribution relates to the finding of the increasing role of the Bay Area in patenting overall.

## **2. Data and empirical strategy**

We use patents granted by the US Patent and Trademark Office (USPTO) as our measure of invention. Because of the delay between patent application and grant date, we date patents using the year of application. We have data on patents granted between 1976 and 2012, and our analysis data set includes patents with application dates between 1976 and 2008. We cut off the last four years of the data because of lags between year granted and year filed. Generally, we start to see a decline in patenting in 2008, suggesting right truncation may be an issue for the last few years of our data. The trends we identify appear long before 2008.

Patents have been shown to provide a useful measure of a firm's intangible stock of knowledge (Hall et al. 2005). Their limitations are well known. Not all patents meet the USPTO criteria for patentability (Jaffe and Trajtenberg 2002). Not all inventors seek to patent, and many use alternative means to appropriate value from their inventions. Further the propensity to patent has changed over time during our sample (e.g., Hall and Ziedonis 2001), this was particularly the case for patents related to software which grew rapidly toward the end of our sample period due to legal changes which strengthened the legal rights of patents in this area (e.g., Graham and Mowery 2003, Hall and MacGarvie 2010). Our use of patent citations as a measure of knowledge flows between successive generations of inventions can also create measurement error (Roach and Cohen 2013). We are comfortable with using patents in this context because our primary focus is on changes in the geographic distribution of patenting within broad technology areas over time. While the propensity to patent has changed across patent classes over time, we do not believe it has changed significantly across geographic locations patenting within a patent class.

We map inventors to counties and MSAs using the zip code of the location of the inventor. We used consolidated MSAs (CMSAs) where those were present. This will be particularly important for our analysis of the Bay Area, which includes several component PMSAs such as Oakland, San Francisco, San Jose, Santa Cruz-Watsonville, Santa Rosa, Stockton-Lodi, Vallejo-Fairfield, and Napa.

For most of the analysis that follows, we do not weight by citations. For multi-author patents, we divide by the number of authors. For example, if a patent has 1 author in the Bay Area and 2 authors in Boston, it would count as 1/3 of a patent in the Bay Area and 2/3 of a patent in Boston. Our results are generally robust, and often stronger, using three year and five year citation-weighted measures. For example, using either three or five year citation weights, the Bay Area surpasses New York City as the location with the most patents three years earlier than with the unweighted measure.

Some of our results require us to recognize a consistent identifier for assignees (especially firms) within a particular application year. For this paper, we do not seek to identify changes in patenting activity within firms over time. Because assignee names are not coded consistently within the patent data, the challenges of mapping patents to assignees is well known. No prior data set provides a complete set of cleaned assignee names during our sample period. Using the data file of standardized names in the NBER database

and the names in ICT industries compiled by Ozcan and Greenstein (2013) as starting points, we cleaned the assignee names to create our own assignee identifier.

Our analysis requires us to identify patents that represent inventions related to ICT, or inventions that draw upon the stock of knowledge related to ICT. As is well known, identifying such inventions through the patent data is notoriously difficult (see, e.g., Graham and Mowery 2003, Bessen and Hunt 2007; Hall and MacGarvie 2010). As a result, we use several different definitions based on the primary class of the patent and explore the robustness of our results to four alternatives. We discuss the construction of these alternatives in the online appendix.

Our data contain a total of 2,213,271 patents. In 1976, there were 41,100 new patents issued from the PTO. At the peak of our data in 2007, there were 100,832 patents.

We present our results at the year level, as aggregated means over the 33 years from 1976 to 2008 inclusive. In particular, our results are presented as graphs of time trends of the fraction of patents each year that meet some criteria such as being based in the Bay Area. This is therefore a descriptive exercise that tests whether the results are consistent with increasing coagglomeration in the San Francisco Bay Area over time. We have not determined the primary cause(s) of the observed patterns.

### **3. Results**

#### **a) Patenting across locations**

Given the overall rise in the propensity to patent, all major cities had an increase in the number of patents. We explore the fraction of all US patents by city, thereby controlling for the overall trend.

Figure 1 shows the increasing importance of the Bay Area as a fraction of US patenting. Figure 1a compares the top 10 cities in the United States, defined by the total number of patents between 1976 and 2008. In 1976, New York City was the dominant center for patenting, with just under 15% of all patents. Los Angeles was second and Chicago was third. Generally, patenting was highly correlated with population. The Bay Area rose steadily as a fraction of patenting in the 1970s and 1980s, and then the trend increased in the 1990s before settling down at the earlier rate of increase in the 2000s. In 1995, the Bay Area surpassed New York City as the US location with the largest number of patents. Figure 1b contrasts the 11<sup>th</sup> through 20<sup>th</sup> cities in patenting with the Bay Area in order to show that no other city has a rise similar in scale.

Figure 1c combines locations into four groups: the Bay Area, New York City, the 18 other cities in the top 20, and all other locations. Generally while New York and locations outside the top 20 are falling as a proportion of patenting, the Bay Area is rising quickly, and the other 18 cities in the top 20 are rising slightly (42.6% in 1976 to 46.1% at the peak level in 2004).

#### **b) Patenting across types of patents**

The Bay Area has had a cluster of ICT firms for many years. Therefore, one reason the Bay Area is becoming an increasing large fraction of patenting is that overall increase in ICT patents. Figure 2 displays this

increase using the Hall, Jaffe, and Trajtenberg (HJT) definitions of patent classes. Computers and Communication (Class 2) went from under 10% of patents to over 30% of patents between 1980 and 2005. Some of this growth may reflect changes in the propensity to patent software and other ICT inventions (e.g., Graham and Mowery 2003, Hall and Ziedonis 2001) that have been encouraged by sympathetic treatment in the courts and the PTO. Drugs and Medical (Class 3) tracked the increase in Computers and Communication until the mid-1990s but then settled back to around 13% of patents.

We offer the first evidence of the coagglomeration hypothesis with Figure 3, which shows the fraction of patents that are in the Bay Area by broad class. The increase is sharpest in Computers and Communication and in Electrical and Electronic (Class 4). It is also visibly noticeable in Chemicals (Class 1), Drugs and Medical, and Mechanical (Class 5). In Other (Class 6) the increase is smaller, rising from 3.8% in 1976 to peak of 6.3% in 2004 before falling back to 4.4% in 2008. Thus, for five of six broad patent classes, we see a noticeable rise in the proportion of patents coming from the Bay Area.

One possibility is that many of the patents in Chemicals, Drugs and Medical, and Mechanical classes are ICT-based. Software has increasingly been used as an input into a wider array of inventions in other patent categories (Arora, Branstetter, and Drev 2013; Branstetter, Drev, and Kwon 2015), taking advantage of increasingly inexpensive and more capable electronics, especially processors. Figure 4 provides alternative measures of ICT and non-ICT patents to account for this possibility, and examines the trend over time.

Panel A of Figure 4 provides the narrowest definition. It defines ICT patents as patents in HJT Class 2: Computers and Communication. The solid line at the top of the graph shows the increasing proportion of Class 2 patents that are in the Bay Area, replicating the solid line at the top of Figure 3. Panel B provides a wider definition, including software patents as defined by Graham and Mowery (2003). Panel C provides a still-wider definition, adding to the definition in panel B all Electrical and Electronic patents (HJT Class 4). Panel D provides our widest definition, which uses the definition in Panel C as a starting point and then widens it to include software patents identified through a keyword search as in Bessen and Hunt (2007) and software patents identified in Graham and Vishnubhakat (2013). This broadest definition includes 74% of all patents in 2008 and is likely to include many false positives. Using all four, increasingly broad, definitions of ICT patents, the solid lines show that the proportion of all ICT patents in the Bay Area has risen sharply since the 1970s.

The dotted line identifies all patents that cite ICT patents. In each panel, the ICT patents are defined as in the previous paragraph. These patents are not explicitly categorized as ICT using the definition above, but they are connected through citation and therefore build on ICT invention. There is a clear trend toward an increasing proportion of these patents in the Bay Area, providing another explanation for the rise of Bay Area patents.

Together, the above suggest the following: ICT is an increasingly large fraction of patents; the Bay Area is an increasingly large (and even dominant) fraction of ICT patents; and the Bay Area is an increasingly large fraction of patents that cite ICT patents. Given prior results on agglomeration of the ICT industry in the Bay Area, perhaps none of these results are surprising, though we believe that the results on geography of patents that cite ICT are not previously documented. These all could result from agglomeration of software invention near the location of the firms producing electronics, computing, and communications.

The evidence for coagglomeration of invention appears in the dashed line in Figure 4: The Bay Area is an increasing fraction of non-ICT US patents, even for the broadest definitions of software. Panel D shows that the fraction of non-ICT patents in the Bay Area rises from 3.9% to 6.2% from 1976 to 2008, a 59% increase. Under the narrower software definition in Panel C, the value rises from 3.9% to 6.9%. Dropping electronics, as in Panel B, the proportion of non-ICT patents in the Bay Area rises from 4.3% to 9.5%.

While these figures are more modest than the increase in ICT patents, they still suggest an increasingly important role for the Bay Area, relative to all other areas, in US non-ICT patenting. Figure 5 compares the Bay Area to the four other top patenting cities in the United States. Panel A uses the narrower ICT definition that includes Software and Computers and Communications patents in Figure 4 panel B. Under this definition, the Bay Area overtook New York as the top location for non-ICT patenting in 2000. Panel B includes Electrical and Electronic patents as in Figure 4 panel C. Under this broad definition, the Bay Area was second behind New York for most of the period from 1997 to 2008. Using the broadest definition (as in Figure 4 panel D) yields a similar pattern (though, as noted above, that definition will include many false positives on software patents).

Overall, we interpret these results to suggest that we cannot reject coagglomeration of invention. The increase in patenting in the Bay Area is not entirely attributable to the increasing fraction of ICT patents in overall patenting.

### c) Patenting across types of patentees

Does the evidence for coagglomeration reflect some other factors, such as firm size? We look at the fraction of patents in the Bay Area by size of patentee. We find that size does not explain the results.

We split patentees into four categories: Independent inventors, firms in the 99<sup>th</sup> percentile of firm patentees, firms between the 50<sup>th</sup> and 99<sup>th</sup> percentile of firm patentees, and firms below the 50<sup>th</sup> percentile of patentees. Figure 6 Panel A shows that the fraction of patents in the Bay Area is rising sharply over time for all four groups. The 99<sup>th</sup> percentile group has a more discontinuous increase, largely driven by a sharp increase in the proportion of top patenting firms that are ICT firms in the 1990s. The only exception to the general trend is that, since 2000, the proportion of all patents from smaller patenting firms (below the 50<sup>th</sup> percentile) in the Bay Area has declined to the levels of the early 1990s. Panel B looks at ICT patents only (defined to include software and electronics as in Figure 4 panel C) and shows a steady rise in the proportion of such patents in the Bay Area between 1976 and 2006, again with the exception of smaller patenting firms since 2000.

Panel C looks at non-ICT patents, again defined as in Figure 4 panel C. The fraction in the Bay Area has been rising steadily for firms below the 99<sup>th</sup> percentile and for independent inventors. In the 99<sup>th</sup> percentile, results are noisier as the entry or exit from the 99<sup>th</sup> percentile of one firm can make a meaningful difference. For example, the value is generally under 3% with the exception of the period 1997 to 2002. During this period, at least one of Genentech, Advanced Cardiovascular Systems, and Incyte Pharmaceuticals was in the 99<sup>th</sup> percentile of non-ICT patentees each year.

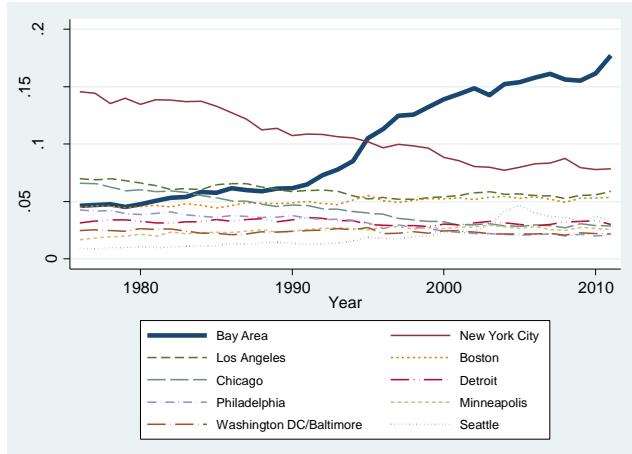
## 4. Conclusions

We have documented an increase of the fraction of US patenting of all kinds that occurs the Bay Area that is disproportionate to population growth and occurs within a variety of patent classes. This partly results from the agglomeration of invention near the production of firms who use the invention, and who themselves agglomerate in one area. We also think it offers evidence of coagglomeration, the clustering of invention from many distinct types of invention into one geographic area.

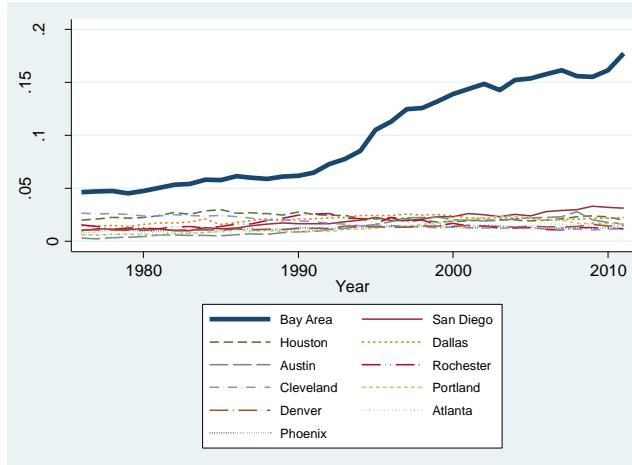
While we do not know the cause of the rise in coagglomeration of many patent types in the Bay Area, our results suggest that any possible explanation must be broad-based. In particular, any explanation must account for growth in the fraction of ICT and non-ICT patents in the Bay Area and for the increase to be true of large inventing firms, small inventing firms, and independent inventors.

**Figure 1: Fraction patents by location**

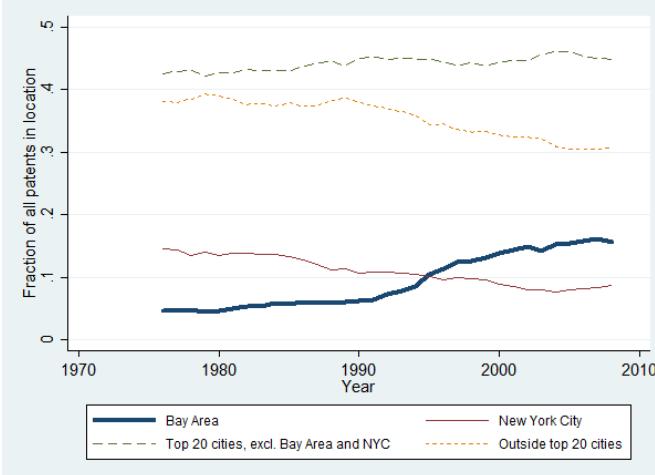
**Panel A: Top 10 cities**



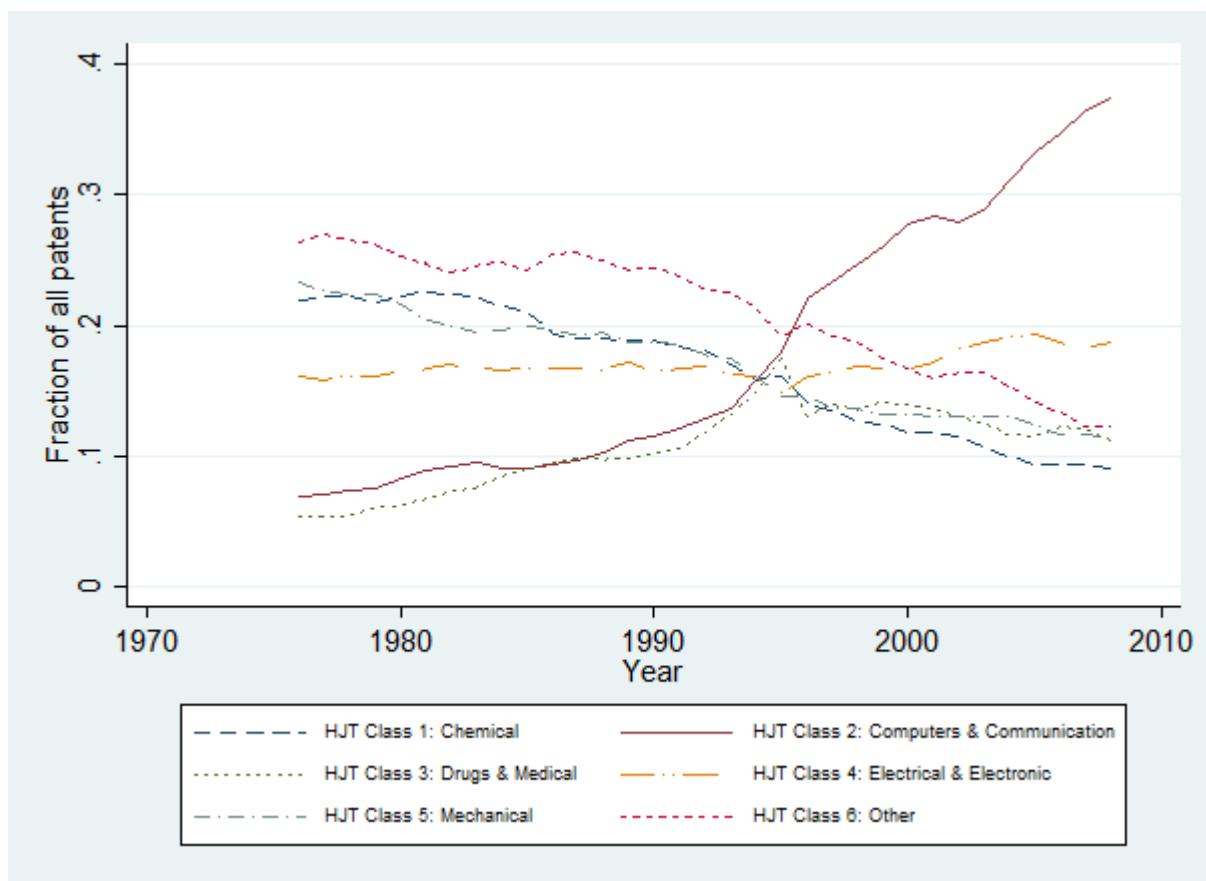
**Panel B: Next 10 cities plus Bay Area**



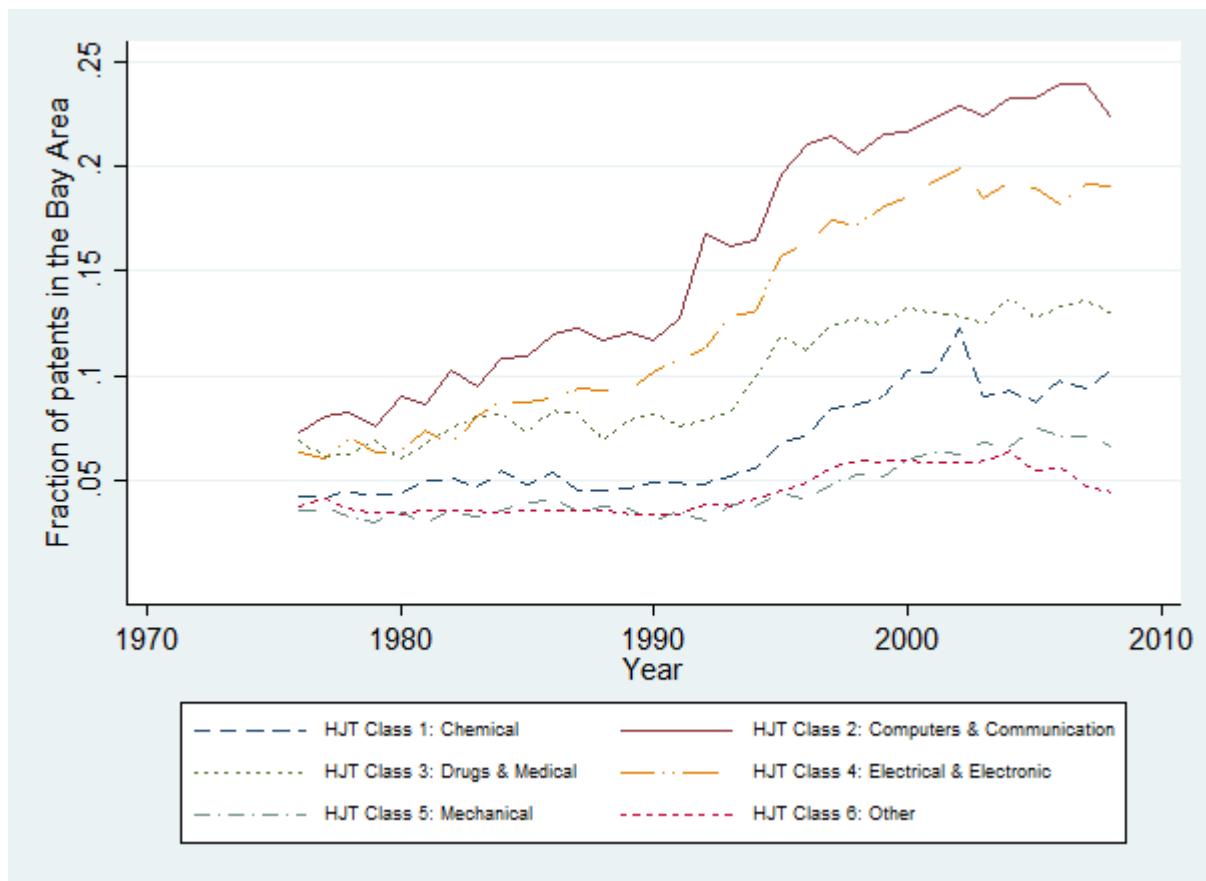
**Panel C: Bay Area, New York City, Other top 20 cities, and elsewhere**



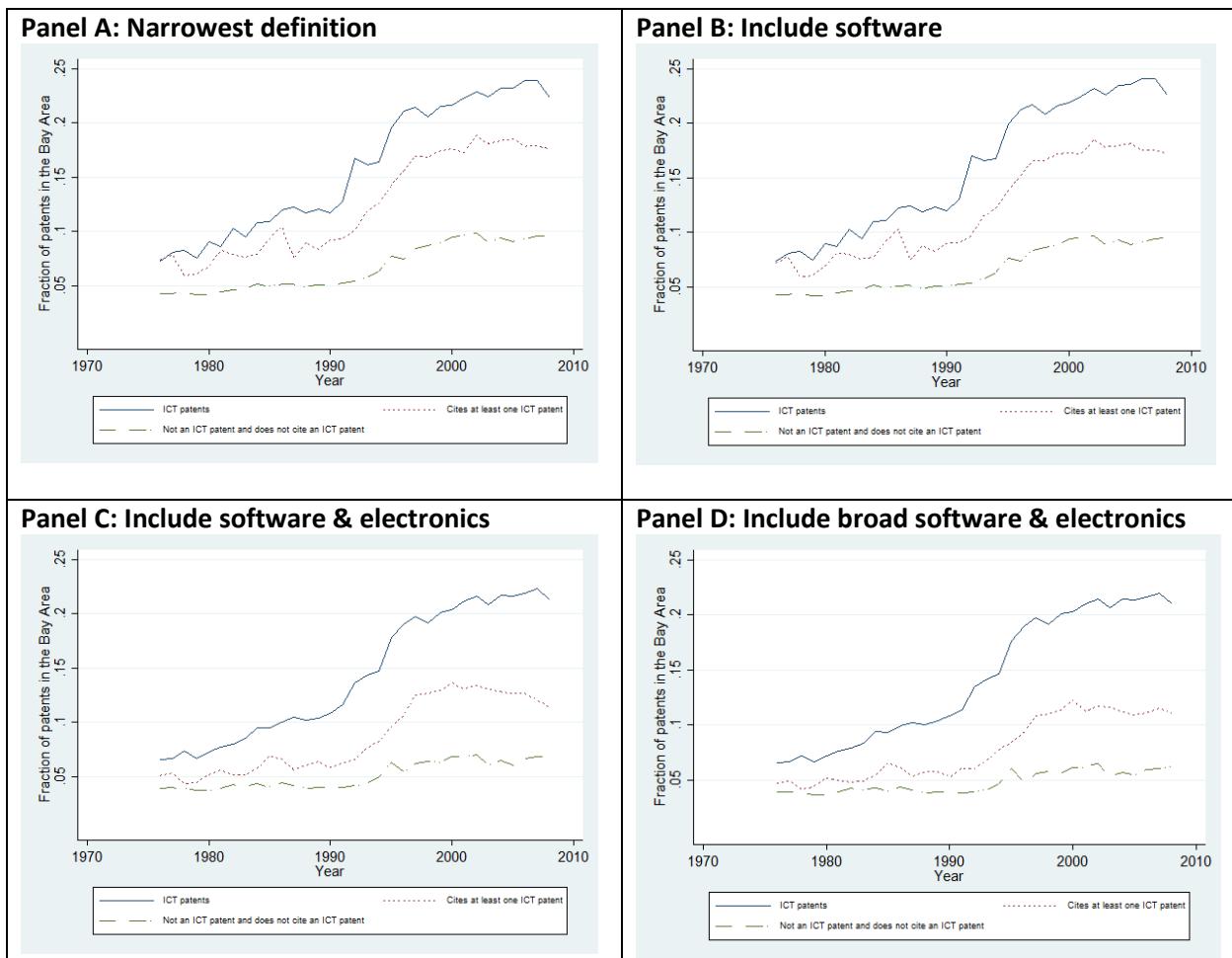
**Figure 2: Information and communication technology is becoming increasingly important**



**Figure 3: Fraction of innovation in the Bay Area by HJT Class**

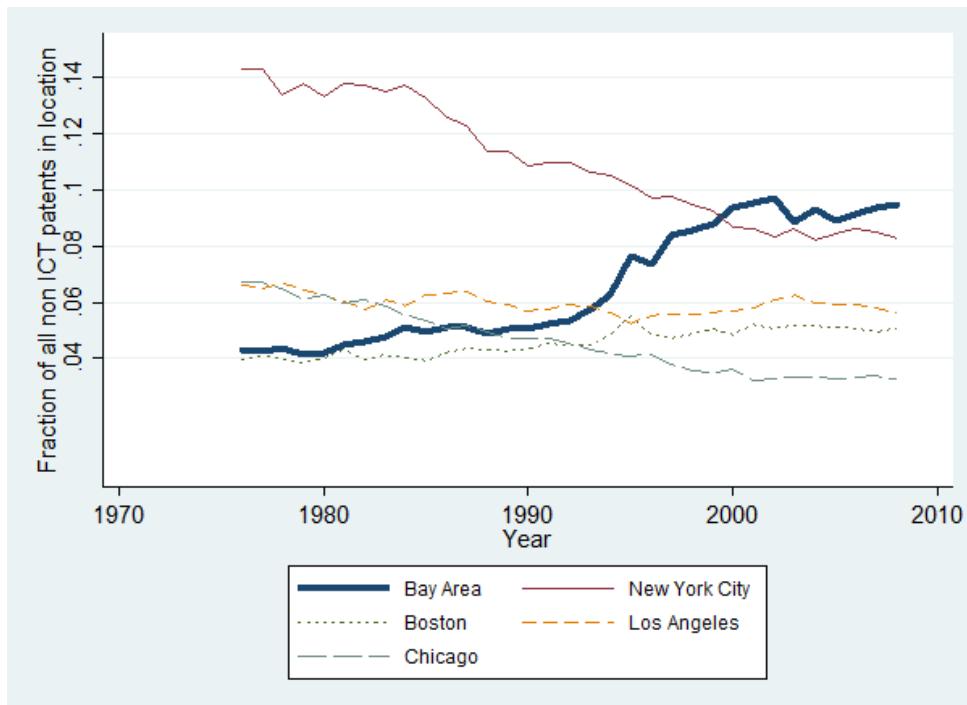


**Figure 4: Alternative definitions of ICT**

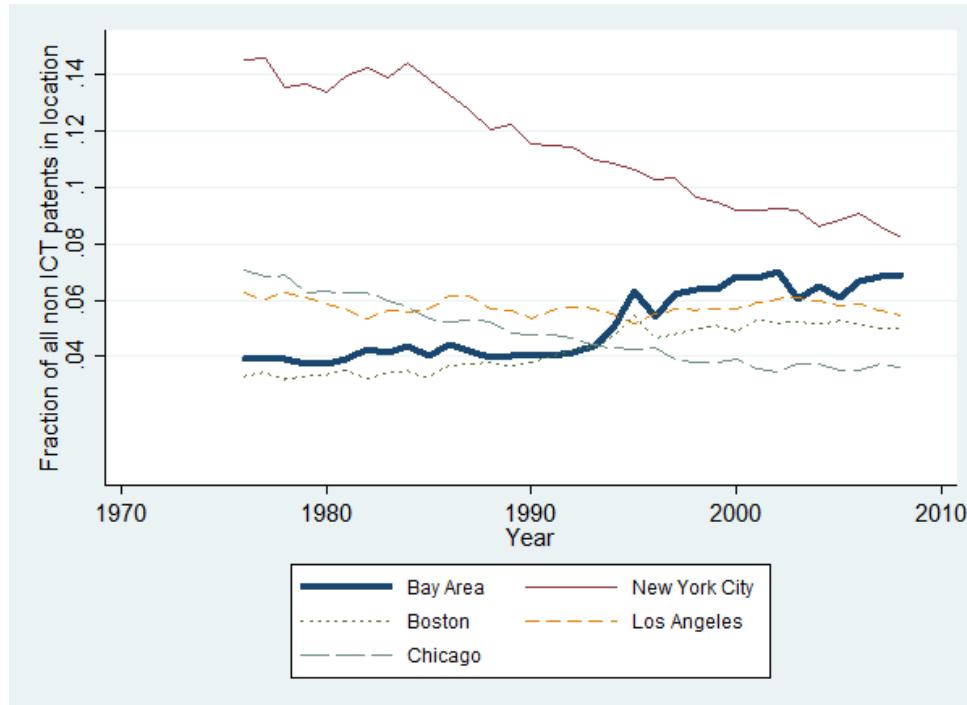


**Figure 5: Alternative definitions of no ICT, comparison across top 5 cities**

**Panel A: No ICT where ICT defined as ICT or main software definition, or cites those**

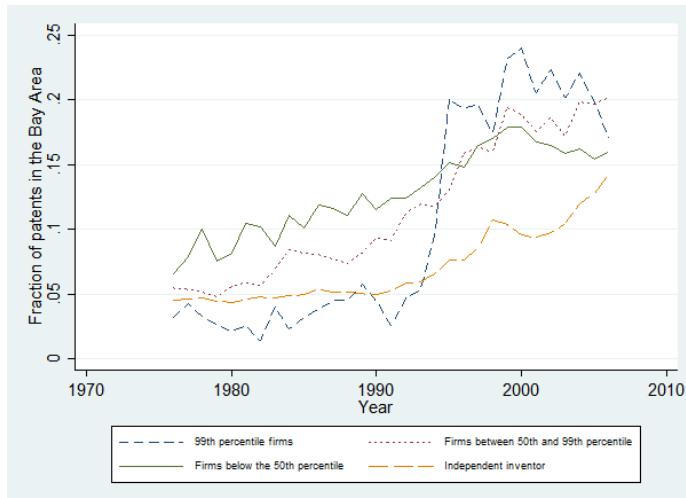


**Panel B. NO ICT where ICT defined as ICT or main software definition or electronics or cites those**

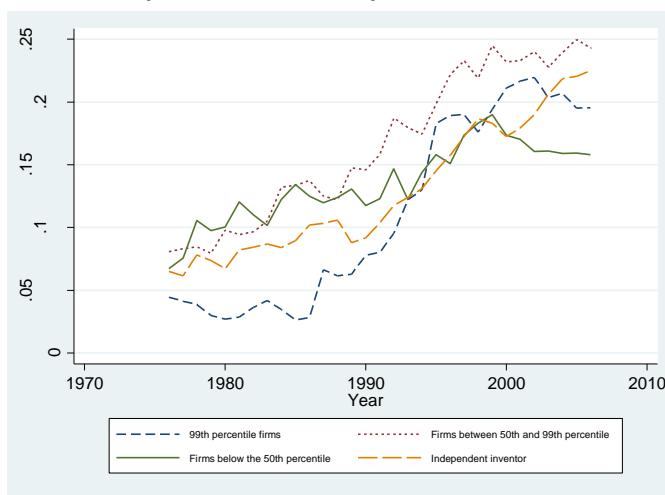


**Figure 6: Bay Area Fraction patenting by type of patentee**

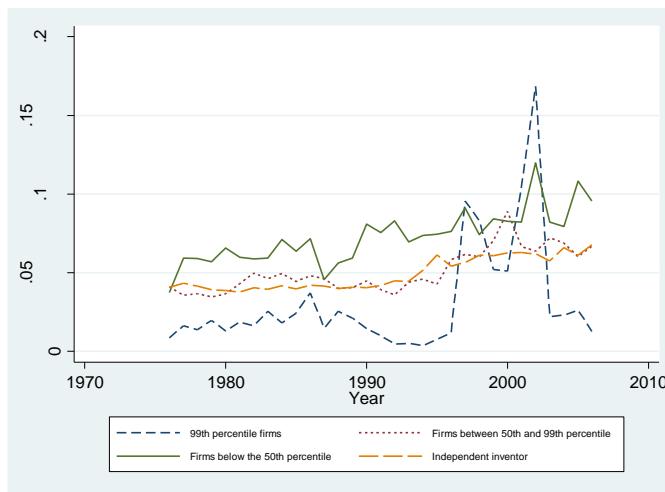
**Panel A: All patents**



**Panel B: ICT patents, defined by best software definition and electronic**



**Panel C: NON-ICT patents, defined by best software definition and electronic**



## Extended Reference List

- Aarland, Kristin, James C. Davis, J. Vernon Henderson, and Yukako Ono. 2007. Spatial organization of firms: the decision to split production and administration. *RAND Journal of Economics* 38(2): 480-494.
- Almeida, Paul and Bruce Kogut. 1999. Localization of Knowledge and the Mobility of Engineers in Regional Networks. *Management Science* 45(7): 905-917.
- Arora, Ashish, Lee G. Branstetter, and Matej Drev. 2013. Going Soft: How The Rise Of Software-Based Innovation Led To The Decline Of Japan's It Industry And The Resurgence Of Silicon Valley. *Review of Economics and Statistics*, 95(3): 757-775.
- Arora, Ashish and Alfonso Gambardella. 2005. *From Underdogs to Tigers: The Rise and Growth of the Software Industry in Brazil, China, Ireland, and Israel*. Oxford, UK: Oxford University Press.
- Audretsch, D., Feldman, M., 1996. R&D spillovers and the geography of innovation and production. *American Economic Review*. 86, 630–640.
- Azoulay, P., J Graff Zivin, and J. Wang. 2010. Superstar extinction. *Quarterly Journal of Economics*. 125 (2), 549–589.
- Barro, Robert J. and Xavier Sala-i-Martin. 1991. "Convergence across States and Regions." *Brookings Papers on Economic Activity*, (1991)1: 107-158.
- Bessen, James, and Robert Hunt. 2007. An Empirical Look at Software Patents. *Journal of Economics and Management Strategy*. 16(1), 157-189.
- Blanchard, Olivier Jean and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, 1992(1): 1-61.
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. Americans Do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review* 102(1): 167-201.
- Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. 2009. "The Distinct Effects of Information Technology and Communication Technology on Firm Organization." NBER Working Paper #14975.
- Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. 2002. Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *Quarterly Journal of Economics* 117(1): 339-376.
- Bresnahan, Timothy and Shane Greenstein. 1996. Technical progress in computing and in the uses of computers *Brookings Papers on Economic Activity: Microeconomics* (1996): 1-78.
- Branstetter, Lee, Matej Drev, and Manho Kwon. 2015. Get with the Program: Software-Driven Innovation in Traditional Manufacturing. Working paper.

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

Bresnahan, Timothy, and Franco Malerba. 1999. Industrial Dynamics and the Evolution of Firms' and Nations' Competitive Capabilities in the World Computer Industry. In David Mowery and Richard Nelson eds, *Sources of Industrial Leadership: Studies of Seven Industries*. Cambridge University Press. Chapter 3.

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

Carlino, Gerald, and William Kerr. 2015. Agglomeration and Innovation. In Gilles Duranton, Vernon Henderson, and William Strange eds. *Handbook of Regional and Urban Economics Volume 5*. p. 349-404.

Charlot, Sylvie and Gilles Duranton. 2006. Cities and workplace communication: Some quantitative French Evidence. *Urban Studies* 43(8): 1365-1394.

Chen, Henry, Paul Gompers, Anna Kovner, and Josh Lerner. 2010. Buy local? The geography of venture capital. *Journal of Urban Economics* 67(1): 90-102.

Chinitz, B., 1961. Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review*. 51 (2), 279–289.

Delgado, M., Porter, M., Stern, S., 2010. Clusters and entrepreneurship. *Journal of Economic Geography*. 10 (4), 495–518

Delgado, Mercedes, Michael Porter, and Scott Stern, 2014. Clusters, Convergence, and Economic Performance. *Research Policy* 43(10), 1785-1799.

Duranton, G., 2007. Urban evolutions: the fast, the slow, and the still. *American Economic Review*. 97, 197–221

Duranton, Gilles and Diego Puga. 2005. From sectoral to functional urban specialization. *Journal of Urban Economics* 57: 343-370.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review*, 100(3): 1195-1213.

Feldman, M., 1994. *The Geography of Innovation*. Kluwer Academic, Boston, MA.

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

- Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2012. The Internet and Local Wages: A Puzzle. *American Economic Review* 102(1): 556-575.
- Forman, Chris and Nicolas van Zeebroeck. 2012. From wires to partners: How the Internet has fostered R&D collaborations among firms. *Management Science* 58(8): 1549-1568.
- Franco, April, and Matthew Mitchell. 2008. Covenants not to Compete: Labor Mobility and Industry Dynamics. *Journal of Economics and Management Strategy* 17(3), 581-606.
- Garcia-Vicente, Florencia, Daniel Garcia-Swartz, and Martin Campbell-Kelly. 2014. The Hydra of Lerna: Economic-Geography Perspectives on America's Early Computer Clusters. Working paper, Warwick University.
- Gaspar, J. and E. L. Glaeser. 1998. Information Technology and the Future of Cities. *Journal of Urban Economics* 43: 136-156.
- Glaeser, Edward. 2005. Urban Colossus: Why is New York America's Largest City? *FRBNY Economic Policy Review*. December. p. 7-24.
- Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer. 1992. "Growth in Cities." *Journal of Political Economy*, 100(6): 1126–52.
- Glaeser, E., Kerr, W., Ponzetto, G., 2010. Clusters of entrepreneurship. *Journal of Urban Economics*. 67 (1), 150–168
- Glaeser, E., Kerr, S., Kerr, W., 2015. Entrepreneurship and urban growth: an empirical assessment with historical mines. *Rev. Econ. Stat.*, forthcoming.
- Glaeser, E.L. and Giacomo A. M. Ponzetto. 2007. Did the Death of Distance Hurt Detroit and Help New York? NBER Working Paper 13710.
- Graham, Stuart, and David Mowery. 2003. Intellectual Property Protection in the U.S. Software Industry. In Wesley Cohen and Stephen Merrill eds, *Patents in the Knowledge Economy*, National Research Council, Washington DC. p. 219-258
- Graham, Stuart, and Saurabh Vishnubhakat. 2013. Of Smart Phone Wars and Software Patents. *Journal of Economic Perspectives* 27(1), 67-86.
- Greenstein, Shane. 2015. *How the Internet Became Commercial: Innovation, Privatization, and the Birth of a New Network*. Princeton University Press; Princeton, NJ.
- Hall, Bronwyn, and Megan MacGarvie. 2010. The private value of software patents. *Research Policy* 39(7), 994-1009.
- Hall, B. H., R. Ziedonis 2001. The Patent Paradox Revisited: An Empirical Study of Patenting in the US Semiconductor Industry, 1979–1995. *RAND Journal of Economics* 32: 101-128.

- Hannigan, Thomas J., Marcelo Cano-Kollmann, and Ram Mudambi. 2015. Thriving innovation amidst manufacturing decline: the Detroit auto cluster and the resilience of local knowledge production. *Industrial and Corporate Change* 24(6): 1-22.
- Henderson, Vernon. 2003. "Marshall's Scale Economies." *Journal of Urban Economics* 53(1): 1-28.
- Hensley, Robert W. and William C. Strange. 2014. Coagglomeration, Clusters, and the Scale and Composition of Cities. *Journal of Political Economy*. 122(5), 1064-1093.
- Higgins, Matthew J., Daniel Levy, and Andrew T. Young. 2006. "Growth and Convergence across the United States: Evidence from County-Level Data." *Review of Economics and Statistics*, 88(4): 671–81.
- Hsieh, Chang-Tai, and Enrico Moretti. 2015. Why Do Cities Matter? Local Growth and Aggregate Growth. NBER Working Paper 21154.
- Isaacson, Walter. 2011. *Steve Jobs*. Simon & Schuster, New York.
- Jaffe, Adam and Manuel Trajtenberg. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. MIT Press, Cambridge, MA.
- Jaffe, A., M. Trajtenberg and R. Henderson 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* 108(3): 577-598.
- Kerr, W., 2010. Breakthrough inventions and migrating clusters of innovation. *Journal of Urban Economics*. 67 (1), 46–60.
- Kerr, William, and Scott Kominers., 2015. Agglomerative forces and cluster shapes. *Review of Economics and Statistics*, forthcoming
- Levy, Frank and Richard J. Murnane. 2005. *The New Division of Labor: How Computers are Creating the Next Job Market*. Princeton: Princeton University Press.
- Macher, J. and D. Mowery. 2008. *Innovation in Global Industries: U.S. Firms Competing in a New World*. Washington, DC: National Academies Press.
- Magrini, Stefano. 2004. "Regional (Di)Convergence." In *Handbook of Regional and Urban Economics*, Volume 4, ed. J. Vernon Henderson and Jacques-Francois Thisse, 2243–92. Amsterdam: Elsevier North-Holland.
- Marshall, Alfred. 1920. *Principles of Economics*. 8th ed. New York: Porcupine Press.
- Marx, M. J. Singh, and L. Fleming, "Regional Disadvantage? Employee Non-compete Agreements and Brain Drain." *Research Policy* 44(2):394-404 (2015).
- Moretti, E., 2004. Workers' education, spillovers and productivity: evidence from plant-level production functions. *American Economic Review*. 94, 656–690.

- Moretti, E., 2012. The New Geography of Jobs. Houghton Mifflin Harcourt, New York. Moretti, E., 2013. Real wage inequality. *American Economic Journal: Applied Economics*. 5 (1), 65–103.
- Moser, P., 2011. Did patents weaken the localization of innovations? Evidence from World's Fairs. *Journal of Economic History*. 71 (2), 363–381.
- Ozcan, Yasin and Shane Greenstein. 2013. The (de)Concentration of Sources of Inventive Ideas: Evidence from ICT Equipment. Working Paper.
- O'Sullivan and Strange. 2015. The emergence of Coagglomeration. Working paper, University of Toronto.
- Perlman, Elizabeth Ruth. 2015. Dense Enough to be Brilliant. Working paper. Boston University.
- Powell, Walter W., and Eric Giannella. 2009. Collective Invention and Inventor Networks. *Handbook of Economics of Innovation*. Bronwyn Hall and Nathan Rosenberg Eds.
- Powell, Walter W., Kelley Packalen, and Kjensten Whittington. 2011. Organization and Institutional Genesis: The Emergence of High-Tech Clusters in the Life Sciences.
- Roach, Michael and Wesley M. Cohen. 2013. Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research. *Management Science* 59(2): 504-525.
- Rosenberg, Nathan, 1963, Technology Change in the Machine Tool Industry, 1840-1910. *The Journal of Economic History*, 23(4), pp. 414-443.
- Rosenthal, S., Strange, W., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.), *Handbook of Regional and Urban Economics*, vol. 4. NorthHolland, Amsterdam, pp. 2119–2171.
- Rosenthal, S., Strange, W., 2008. The attenuation of human capital spillovers. *Journal of Urban Economics*. 64 (2), 373–389.
- Saxenian, A., 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, Cambridge, MA.
- Storper, M. and Venables, A. J. 2004. Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*. 4(4): 351–370.
- Thompson, P., Fox-Kean, M., 2005. Patent citations and the geography of knowledge spillovers: a reassessment. *American Economic Review* 95 (1), 450–460
- von Hippel, Eric. 1998. Economics of Product Development by Users: The Impact of “Sticky” Local Information. *Management Science* 44(5): 629-644.
- Waldinger, F., 2012. Peer effects in science: evidence from the dismissal of scientists in Nazi Germany. *Rev. Econ. Stud.* 79, 838–861.

Watzinger, Martin. 2015. Antitrust, Patents, and Cumulative Innovation: Evidence from Bell Labs. Working paper.

Zook, Matthew A. 2005. *The Geography of the Internet Industry: Venture Capital, Dot-coms, and Local Knowledge*. Blackwell Publishers.

## **Online Appendix: Identification of patents related to information and communication technologies (ICT)**

As noted in the paper, we take four approaches to identify inventions that are related to ICT, or that draw upon the stock of knowledge related to ICT. All four approaches are based on the primary class of the patent.

Our first approach uses the categories from the NBER Patent Data project's 2006 update of the categories developed by Hall, Jaffe, and Trajtenberg (HJT) (2001). We identify ICT patents as HJT category 2, which includes subcategories such as computer hardware & software, communications, computer peripherals, information storage, and electronic business methods and software, among others.

Our second approach adds patents specifically identified as software patents to the set defined in our first approach. It relies upon identifying software-related patent classes in the International Patent Classification (IPC) system. Specifically, we use the approach defined in Graham and Mowery (2003), which identified software patents as being in the class/subclasses "Electric Digital Data Processing" (G06F), "Recognition of Data; Presentation of Data; Record Carriers; Handling Record Carriers" (G06K), and "Electric Communication Technique" (H04L). Compared to USPTO classes, software is more concentrated in the IPC system. The Graham-Mowery approach has been used as inputs into the definitions used in a variety of papers that have required identification of software patents, including Hall and MacGarvie (2010) and Arora, Branstetter, and Drev (2013), among others.

Our third definition adds patents in HJT category 4 (electronics) to the set of ICT patents defined in our second definition. This definition will capture electronics inventions that are closely related to ICT, such as inventions related to semiconductor devices.

Our last approach widens the set of ICT patents defined in approach 3 in two ways. First, it adds patents identified as software through a keyword search. Specifically, we identify software patents using the keywords from Bessen and Hunt (2007) to search for specific software-related strings in the titles and abstracts of patents in our sample. This approach also includes patents from the USPTO classes identified by Graham and Vishnubhakat (2013) as containing general purpose software or software that is specific to some type of hardware.

## **References**

- Arora, Ashish, Lee G. Branstetter, and Matej Drev. 2013. Going Soft: How The Rise Of Software-Based Innovation Led To The Decline Of Japan's IT Industry And The Resurgence Of Silicon Valley. *Review of Economics and Statistics*, 95(3): 757–775.
- Bessen, James, and Robert Hunt. 2007. An Empirical Look at Software Patents. *Journal of Economics and Management Strategy*. 16(1), 157-189.
- Graham, Stuart, and Saurabh Vishnubhakat. 2013. Of Smart Phone Wars and Software Patents. *Journal of Economic Perspectives* 27(1), 67-86.
- Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg. 2001. The NBER patent citations data file: Lessons, insights, and methodological tools. NBER Working Paper 8498.
- Hall, Bronwyn, and Megan MacGarvie. 2010. The private value of software patents. *Research Policy* 39(7), 994-1009.