

The Geography of Innovation: Patents in Early America

Elisabeth Ruth Perlman

January 2, 2014

This project explores the relationship between urbanization and innovation in nineteenth-century America, as measured through number of patents issued. It seeks to understand how information networks affect innovation by examining three aspects of the geography of innovation: First, it will map the relationship between patents and population for all counties in the United States, not just populous or innovative locations, asking whether a minimum population size or density creates innovative agglomerations. Second, it will test the Sokoloff (1988) hypothesis that increasing market access through the spread of transportation infrastructure leads to an acceleration of innovation. Third, it will use a full-text searchable database of all patents to explore the generation and spread of interrelated ideas across time and space.

The philosophical idea behind innovative agglomerations is that knowledge sharing is social, and for social connections to productively transmit knowledge, people need shared skills, including a vocabulary for discussing innovation. The project will explore how “information in the air,” as Alfred Marshall expressed it, of urban agglomerations interacts with concrete, direct channels of information transmission, such as canals, railroads, and telegraphs, and how population size and density interact with infrastructural improvements in information flows over time.

The shape of the relationship between innovation and population is not well understood. It is clear that large, dense cities are more innovative than other places, but it is also clear that such a population is not enough to induce technological improvements (Kowloon Walled City was not reputed for its great innovations). Many other factors, such as education, availability of credit, and market access are hypothesized as alternative or complimentary causes. As population density increases patenting does not necessarily increase—in my data, described below, while all low density locations do not patent, many medium density places patent and many do not. Only the largest cities, with the highest densities, seem to uniformly always patent. This project examines the relationship between patenting and population at all scales. The patent data is also easily linked with data on other factors, such as Matthew Jaremski’s data on banking or the transportation data discussed below, allowing the examination of a variety of alternative hypotheses about the geography of innovation.

Sokoloff (1988) notes that, between 1805 and 1835 in the eastern United States, counties along the newly-built canals, particularly the Erie, saw a sharp increase in patenting activity, hypothesizing that increased market access led to increased patenting activity. I have constructed a dataset, including the data used in Sokoloff (1988), linking all patents issued between 1790 and 1836 to the counties in which the named inventors reside. With the greater time granularity my data provide, it appears that the increase in patenting activity happened before the arrival of the canal, which is inconsistent with the Sokoloff hypothesis. Linking the patent data created by Tom Nicholas to Jeremy Atack’s transportation data, it also appears that greater patenting activity preceded rather than followed the railroads. Finding the direction of causality among infrastructure, market access, and innovation is critical both for economic history and modelling innovation in an endogenous growth framework.

The time and location of patents already begins to answer these questions, but much more can be gleaned from the full text. Structuring and word usage can reveal underlying flows of information, as seen in the way people write about ideas and technology. Data available on Google books can

be used to see where n-grams appear in general discourse. I have assembled a database of the text of all patents issued 1836-1897, created by optical character recognition, which can be linked to the geographic databases mentioned above. Much as Packalen and Bhattacharya (2012) looks at the use of new phrases in patents 1920-2010 to get a sense of what fields are newly important, location centers of innovation will be revealed by the location of n-grams that appear in patents. Work following Jaffe et al. (1993) has doubted the generality of its conclusion that patents in the same city are far more likely to cite each other. This linguistic information may reveal how connected disparate places are to the same information, and how information about innovation transmits spatially. While it is presumed that increasing the flow of information between places decreases the advantage of being in any one location, it may also be that greater information access increases the advantage of those centrally located in the network.

The full text of patents can also be used to explore details of the innovation that is happening in any one location. Questions about what sorts of innovation cluster around the location of innovations that are general-purpose or foundational can be addressed. I am particularly interested in the location of agriculture vs. non-agriculture patents; whether urban areas produce only non-agriculture innovations.

These patent databases can also be used to explore the relationship between industrial agglomerations, innovation, and long-run viability of cities. Areas highly concentrated in one industry make productivity gains through agglomeration forces, but are also vulnerable to industry-specific shocks, which might lead to a troubling relationship between productivity and robustness. Consider the current plight of Pittsburg or Detroit: innovation concentrated in one sector may make a city vulnerable to the sector's fluctuations. This data can be used to identify which areas are persistently innovative, and which are only innovative in specific sectors, or during particular times.

This project links together several data sources. Congress did its best to compile a list of all patents issued to that date in 1874, the data about patents issued 1790-1836 (which were lost in a fire) has since been updated by many volunteers. I merged the town and county information in the Directory of American Tool and Machinery Patents dataset for this period with a database of historical town names from the AniMap 3.0.2 County Boundary Historical Atlas giving me latitude and longitude coordinates that are then merged with the National Historical Geographic Information System shapefiles of U.S. county boundaries. This allows patent counts by county to be created. The rest of the patent data comes from Tom Nicholas' dataset of patents issued from 1836-1900, which has latitude and longitude coordinates of the listed places on these patents. The work done by Richard Hornbeck helps Michael Haines' US Census Data merge across many years. Jeremy Atack's Transportation Data can also be linked with shapefiles of U.S. county boundaries as above. The textual data was scraped from Westlaw, and can be linked to the other patent data.

I am requesting funds for several purposes: First, to support me in the last year of my PhD program and allow me to concentrate on furthering this research. Second, to hire a research assistant to help resolve the many errors in the 1790-1836 patent data, primarily transcription errors and missing data. In many cases a town is spelled wrong, or a patent's state is misidentified due to transcription error. In addition, when a state has multiple similarly-named towns, which happens surprisingly often, research using a subscription to ancestry.com (\$149 for six months) will help locate inventors. This type of work cannot be automated, and must be carried out by someone proficient at data work.

References

- Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–98.
- Packalen, M. and Bhattacharya, J. (2012). Words in Patents: Research Inputs and the Value of Innovativeness in Invention.
- Sokoloff, K. L. (1988). Inventive Activity in Early Industrial America: Evidence From Patent Records, 1790–1846. *The Journal of Economic History*, 48(04):813–850.