

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

# Fuel of Interest and Fire of Genius: Essays on the Economic History of Innovation Executive Summary

Michael J. Andrews  
*University of Iowa*  
*michael-j-andrews@uiowa.edu*

August 21, 2017

Innovation is the key driver of long-term economic growth. Despite its importance, surprisingly little is known about the identities of inventors, how they create new ideas, or how policy can promote invention. Each chapter of my dissertation uses historical data to answer these questions.

In Chapter 1, I examine the establishment of new colleges throughout U.S. history to better understand the role of institutions of higher education in promoting invention. Establishing a new college leads to a roughly one-third increase in invention. But this increase tends not to come from a college's faculty or graduates; instead, it appears to be driven by the individuals that are attracted to the new college and move to live near it.

In Chapter 2, I attempt to understand the role of informal social interactions in generating new ideas. To do so, I study alcohol prohibition in U.S. states. Prohibition represents an event in which the state imposes a restriction on people's ability to informally interact. This imposition sharply reduces invention in the years immediately following prohibition, before rebounding as people find new places to interact.

In Chapter 3, I examine how invention by different demographic groups has changed over time. Not surprisingly, groups that have been historically marginalized, namely blacks and women, are underrepresented as inventors. What is perhaps more surprising is that the degree of underrepresentation has been very persistent over three quarters of a century.

The results in each of these first three chapters are only as credible as the underlying inventor data used. In Chapter 4, I discuss various patent datasets that are available and the strengths and weaknesses of each.

All of the research in this thesis makes use of *historical* patent data to learn about the identities of inventors and how we can promote the creation of new ideas. Using historical data has a number of advantages. Most obviously, historical data allows for the study of long run effects and provides a perspective missing when using exclusively recent data. Just as importantly, history provides a much larger laboratory in which empirical researchers can explore natural experiments. I exploit two such experiments in Chapters 1 and 2. Finally, in many cases, historical data provides information that is not available in contemporary data

for confidentiality reasons. As a concrete example, the microdata for each U.S. decennial census is sealed for 72 years. The 100% census manuscripts are therefore only available for the decennial censuses in 1940 or earlier. The focus of Chapter 3 is on merging individual names from the patent record to these decennial censuses. Without access to this census information, it is very difficult to learn anything about the demographics of inventors. Thus, in spite of their importance for economic growth, the median inventor is currently an anonymous figure, shrouded in mystery. The work in this thesis helps to remove that shroud of mystery, presenting compelling new facts about who invents and how invention can be encouraged or impeded.

I next describe each of the four chapters in more detail.

## **Chapter 1:**

### **The Role of Universities in Local Invention: Evidence from the Establishment of U.S. Colleges**

Throughout the nineteenth to the mid twentieth century, many new colleges were established in the U.S. Unfortunately, learning how these new colleges contributed to American innovation has proven difficult. The challenge is that colleges are not located randomly: areas that receive new colleges may have many other attractive attributes that make them prime innovation hubs and also attracted colleges in the first place. Conversely, many U.S. land grant colleges were intentionally built in less developed areas in an attempt to bring higher education to those who did not previously have access. Comparing these places to more economically developed regions may understate the true effect of establishing the college.

To overcome these challenges, I exploit historical natural experiments to estimate the true causal effect of establishing a new college on local invention. I am also able to shed some light onto why colleges affect invention in the ways that they do. To identify this causal effect, I consult narrative college histories and compile data on the site selection decisions for a subset of new colleges. With this, I can identify “losing finalist” locations that were strongly considered to become the site of a new college but were not chosen for a nearly-random reason.

A concrete example of a college site selection experiment is useful to understand the empirical approach I use. Georgia Tech was founded in Atlanta in 1886. At the time, prominent citizens in Georgia wanted a technical college, but there was no consensus about where to put it. A number of cities vied to receive the new school. Two of the main rival sites were Atlanta and Macon. Both were known primarily as railway depots located in the interior of the state; the two cities also looked similar in terms of population, demographics, industrial composition, and other dimensions. In October 1886, a site selection committee assembled to vote on the location of the college. For the first 23 ballots, neither Atlanta nor Macon obtained the requisite majority of votes. Finally, on the 24th ballot, Atlanta won over Macon by one vote. It is thus easy to believe that Georgia Tech University could have been located in Macon instead of Atlanta.

In an even more striking case, representatives from North Dakota towns literally drew lots to determine the location of the University of North Dakota and North Dakota State

University. Grand Forks won the state university, while Fargo won the state's land grant college, now known as North Dakota State University. The cases of Georgia Tech and the North Dakota universities were not isolated incidents: while the decisions were occasionally less dramatic, these kinds of college site selection experiments occurred all across the United States, in large towns and small, during the second half of the nineteenth century and first half of the twentieth. I identify 73 cases in which the location of a new college was as good as randomly assigned between the winning county and the losing finalists.

Figure 1 graphically shows the effect of establishing a new college on local invention. I plot logged patenting in the college and losing finalist counties. Year 0 is the year in which a new college is established; everything to the left of 0 represents pre-college years, while everything to the right represents post-experiment years. As can be seen, the college and losing finalist counties patent very similarly before the establishment of a new college. After a new college is founded, however, they begin to diverge. I find that the establishment of a new college caused 33% more patents per year in college counties relative to the losing finalists. Exploring this treatment effect further, I find that the difference between college counties and losing finalists tends to increase over time, and the benefits of colleges appear to spill over and increase patenting in neighboring counties relative to far away counties in the same state.

In the figure, I also plot the “non-experimental counties.” These are counties that are neither the college counties nor the losing finalists. This graph shows clearly that simply comparing college counties to these non-experimental counties, as has often been done in previous research, gives a misleading view of the true effect of establishing a college.

New colleges can increase local patenting in many ways. Most obviously, colleges promote the accumulation of human capital in their students; more educated individuals are more likely to become inventors. Knowledge from colleges may also “spill over” and make people living nearby more inventive. Finally, colleges may increase patenting through a “recruitment” effect, attracting more inventive individuals to live close to the college.

To determine the relative strengths of these effects, I use a novel dataset of college yearbooks and find that a college's alumni and faculty account for less than 8% of the patents granted in that college's county. To be clear, this result does *not* say that college graduates don't patent; but if they do patent, they do so in places other than the site of their alma maters.

I also find that knowledge spillovers to individuals unaffiliated with the college are very small. Instead, the recruitment effect appears to be the predominant channel by which establishing a college increases local patenting. Changes in county population are the primary channel by which colleges affect local invention, accounting for 50-65% of the increase in patenting in college counties relative to the losing finalists.

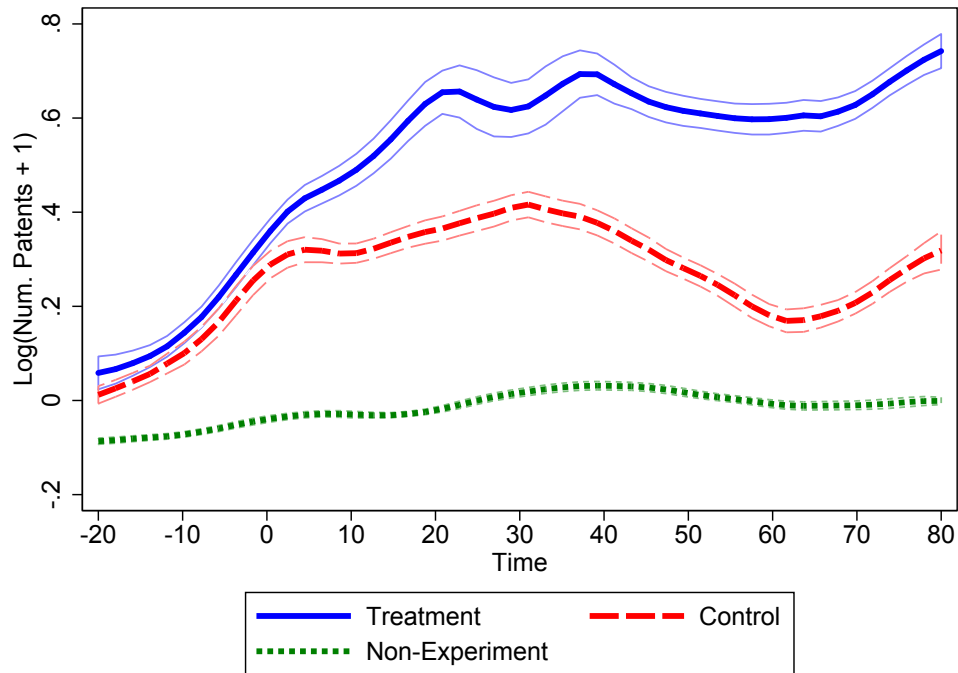


Figure 1: Mean patenting in treatment and control counties after controlling for year effects. The x-axis shows the number of years since the college experiment. The year of the college experiment is normalized to year 0. Everything left of year 0 shows pre-treatment means; everything to the right shows post-treatment means. The y-axis shows smoothed  $\log(\text{Patents} + 1)$ . The smoothed patenting is constructed by regressing  $\log(\text{Patents} + 1)$  on year effects and then plotting the residuals using local mean smoothing with an Epanechnikov kernel function. The treatment counties are represented by the blue solid line. The control counties are represented by the red long-dashed line. The non-experimental counties are represented by the green short-dashed line. Data are for high quality experiments only.

## Chapter 2:

# Bar Talk: Informal Social Interactions, Alcohol Prohibition, and Invention

In the second chapter, I attempt to understand the importance of informal social interactions for the creation of new ideas. Many scholars have argued that informal social interactions, that is, people serendipitously bumping into one another and exchanging ideas, are vital for invention. Unfortunately, it is very difficult to get a sense of the magnitude of the importance of these informal interactions. This difficulty arises for a number of reasons. First, people choose who they talk to. This means that just observing that an invention occurred after a conversation took place does not mean that the conversation caused the invention. Second, estimating the role of social interactions typically requires a great deal of detailed data about peoples' social networks. This data is not available in most contexts.

To overcome these challenges, I study a unique natural experiment in which the state shut down one channel through which people can informally socially interact: alcohol prohibition. More specifically, I examine the effects of state-level alcohol prohibition in the United States. Prior to the enactment of national prohibition in 1920, the U.S. operated under the "local option" doctrine which, as the name implies, meant that counties had the right to determine their own alcohol policies. States also frequently passed statewide alcohol prohibition measures that superseded county-level laws. I compare counties that wanted to remain wet to counties that had decided to go dry and see how each changes before and after the imposition of state-level prohibition. The key idea is that state-level prohibition differentially treats counties depending on their local laws before prohibition.

Figure 2 shows the difference in patenting between wet and dry counties before and after state level prohibition. Year 0 is the first year in which state-level prohibition goes into effect, with all years to the left being pre-prohibition years and to the right being post-prohibition years. The difference between wet and dry counties is stable until the passage of prohibition, at which point patenting in the previously wet counties falls by about 15% relative to the dry counties. As the figure shows, this drop is precipitous in the years immediately following prohibition, but largely rebounds within five years. This is consistent with the drop being caused by a disruption of informal social interactions that dissipates as people find other venues in which they can interact.

I further show that the decline in patenting is larger for men than for women. This is also consistent with the decrease being driven by a disruption of informal social interactions, since women typically did not frequent saloons and so would have been less directly affected by prohibition. Finally, I show that this result is not driven simply by a negative shock to the brewing and distilling industries: even after removing all alcohol-related patents, the decrease in patenting is still present and is nearly the same magnitude.

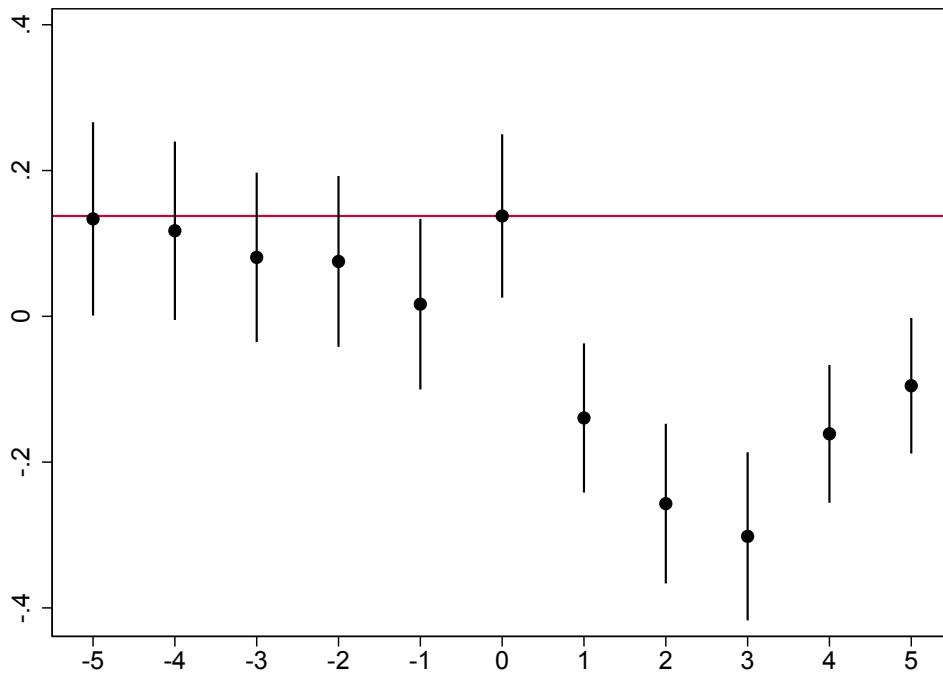


Figure 2: Estimates of the effect of state-level alcohol prohibition on patenting. The x-axis shows the number of years since the enactment of state-level prohibition. The year in which state-level prohibition is enacted is normalized to year 0. Everything left of year 0 shows pre-prohibition means; everything to the right shows post-prohibition means. The y-axis plots logged patenting, so it shows the percentage of additional patents in wet counties relative to dry counties.

## Chapter 3:

# Historical Changes in the Demographics of Inventors in the United States

In this chapter, my coauthors and I attempt to answer the fundamental question of who inventors are. We match data from the Annual Reports of the Commissioner of Patents to the 100% U.S. Population Censuses for 1870, 1880, 1900, 1910, 1920, 1930, and 1940. (Unfortunately, the 1890 census was destroyed in a fire. As mentioned above, all censuses after 1940 are still confidential and cannot be used by researchers.) By matching the patent records to the census, we learn a great deal about who these inventors are. Most notably, we can observe inventors' race, gender, age, and mobility. We are also interested in how these demographic characteristics have changed over time.

Figure 3 plots time series for these demographic characteristics over the seven decades for which we have data. We show whether each group is over or underrepresented relative to its share of the population. A value of one indicates that each group accounts for the same share of patents as its share of the overall population. Or, in other words, a value of one would mean that each patentee can be thought of as just a random draw from the overall population in their county. Values less than one mean that a group is under-represented in the patent record. Likewise, values more than one mean that a group is over-represented in the patent record.

Matching to the census can be challenging because many people share the same name, so that a single patent may match to several people in the census. In the figures, we handle this in several different ways: plotting the maximum and minimum share of patentees with a given characteristics, calculating the "average" characteristic for all patentees with a given name, and plotting only the patentees that were matched uniquely to the census. Each matching technique gives broadly the same results.

These figures show several things. First, blacks and women are consistently underrepresented. Moreover, and more surprisingly, there is very little discernible positive trend in the representation rates for these groups. An alternative method that predicts the race and gender of each patentee by using first names data from the census finds even less of a positive trend than is shown in these figures. We also find that inventors tend to be older than the overall population (we compare patentees to adults aged 18-80 living in the same county) and are more likely to be living in a state other than their state of birth. Both of these patterns are also remarkably stable over time.

The persistence of the trends is particularly noteworthy given the large changes in the political, economic, and social environment over the seventy years we study. We attempt to identify characteristics of the local environment that are correlated with patenting by different groups. Surprisingly, no obvious economic or demographic variables do a good job of explaining patenting by blacks or women. We also explicitly investigate some institutional changes that might be expected to affect patenting for particular groups. The establishment of historically black colleges and universities (HBCUs) appears to increase the number of patents by blacks, but the full magnitude of the increase can be explained by increases in the black population living close to these HBCUs. We also find that the passage of women's suffrage laws at the state level, representing a dramatic and large scale change in the political



and social environment, had no measurable effect on female patenting. There is suggestive evidence that states that adopted women’s suffrage earlier, and thus were likely more open and accepting, saw higher rates of female patenting than did late adopters, but the difference is not statistically significant.

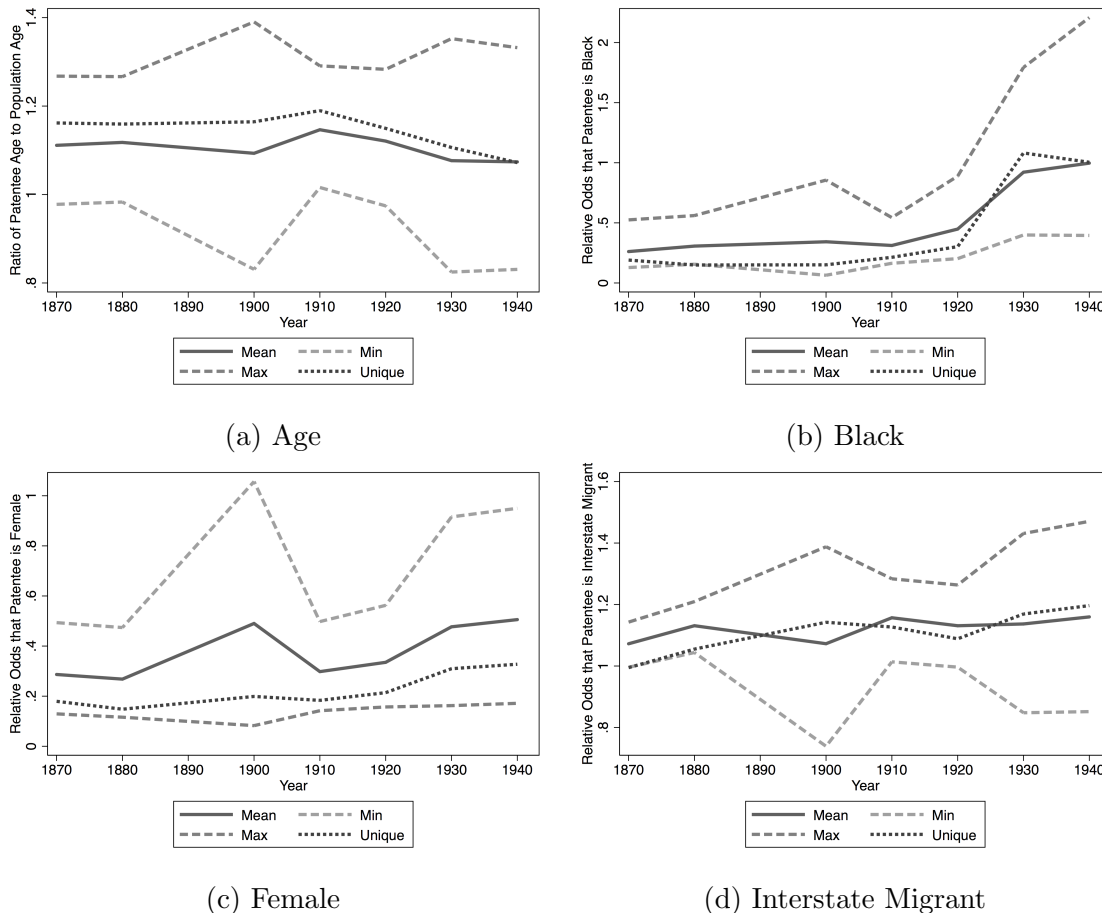


Figure 3: Demographics of patentees relative to county of residence. The mean is the average characteristic across all possible matches for a given patentee, then averaged across all patentees. The max (resp. min) is the “maximum” (resp. “minimum”) value of a particular characteristic across all possible matches, then averaged across all inventors. We also report the values for the set of patentees where we have a unique match. We then report the ratio of these statistics to the average characteristics of the population in the county.

## Chapter 4: Comparing Historical Patent Datasets

As noted above, each of the preceding results is only as credible as the underlying historical patent data that I use. In the fourth chapter, I examine the quality of various different historical patent datasets. I have three main objectives here. First, I aim to inform researchers of the variety of data sources available to study the history of invention in the United States.

Far from being a barren wasteland where data is hard to come by, several research groups have compiled usable datasets that attempt to cover the universe of historical patents. The completeness of these datasets, while not perfect, is comparable to that of datasets covering more recent patenting activity. Furthermore, several resources are available that study particular aspects of historical invention in more detail. A comprehensive understanding of what can be accomplished with each existing dataset prevents researchers from “reinventing the wheel” and building historical datasets from scratch, duplicating the efforts of previous researchers.

Second, I discuss the strengths and weaknesses of each of these datasets. While the creators of each dataset have striven to record the universe of historical U.S. patents, each is constructed from a different source of raw patent data and thus may be of varying quality. Moreover, different authors use different techniques to clean and process the raw patent data, introducing further irregularities across datasets. Finally, the different datasets contain different information about each patent. Understanding how the datasets differ is useful to help researchers decide which is best for a particular application. For example, some datasets contain detailed information on patent classifications, but no information on who the inventors of a particular patent are or where they lived. Understanding what information each dataset contains and how they can be used to complement one another provides researchers with the ability to answer far more questions than would be possible using each dataset in isolation. Just as importantly, even when information may be nominally available, it may be of varying quality across datasets. Learning for which years and patent information a particular dataset should be avoided is also a valuable contribution.

Third and finally, I hope to convey an optimistic message: while the datasets have some differences, overall they paint a remarkably consistent picture of invention in the historical United States. While I focus on the ways in which the datasets deviate from one another, this should not take away from the fact that, when looking at the broad outline of American invention, our conclusions do not depend on the particulars of the dataset that is used. This should provide confidence in the results from earlier research making use of the diverse data available.