

This work is on a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license, <https://creativecommons.org/licenses/by-nc-nd/4.0/>. 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 and telling us what having access to this work means to you and why it's important to you. Thank you.

Changes in the Demographics of American Inventors, 1870-1940*

Sarada

University of Wisconsin, Madison

sarada@wisc.edu

Michael J. Andrews

NBER

mandrews@nber.org

Nicolas L. Ziebarth

Auburn University and NBER

nicolas.lehmannziebarth@gmail.com

May 30, 2019

Abstract

We assemble a novel dataset linking inventors listed in the *Annual Reports of the Commissioner of Patents* to Population Census records spanning 1870 to 1940. We find that inventors are not a random subset of the population. They differ in some unsurprising ways in that they tend to be older, whiter, and more likely male. However, these patterns do change over time. The odds ratio relative to the population as a whole of female inventors increases from a low of 0.07 in 1880 to a high of 0.13 in 1940 and that of non-whites ranges from 0.16 in 1880 to 0.34 in 1940. Both populations remain severely underrepresented throughout the timeframe. We find changes in the occupations of inventors with trends away from farming and towards white collar occupations. We also show the increasing importance of foreign born people in patenting. In 1870, the odds of a foreign born person patenting relative to the population as a whole is nearly 1 and increases to over 1.6 by 1940.

Keywords: Innovation; Demographic Trends; Data Linkage; Patents

JEL Classification: J11, N00, O31

*We thank the Minnesota Population Center for making available the 100% Census files; the Kauffman Foundation, Balzan Foundation, WARF and Northwestern Center for Economic History for financial support; and Siha Lee for excellent research assistance. We also thank Deepak Hegde, Mike Roach, and seminar audiences at the Iowa Macro Workshop, Yale Economic History, Wisconsin School of Business, Wisconsin Agricultural Economics Department, ESSEC (Singapore), Northwestern Searle Conference, Kauffman Foundation, and USPTO for helpful comments.

1 Introduction

Innovation and technological change are at the root of economic growth. While there is rich theoretical literature on innovation and its aggregate effects for growth, much less is known about the individuals who make up these aggregates and generate these breakthroughs. Understanding who these people are is important for understanding who benefits from innovation, and whether certain demographic groups have been systematically precluded from participating in this process. In 1886, Josephine G. Cochran of Shelbyville, Illinois was granted a patent for a dish washing machine. She was trying to make more efficient a task that concerned her on a daily basis. A woman in 1886 being granted a patent (as we will show) was extraordinary for the time and reflects the fact that the set of individuals who invent is not a representative sample of the population. This, in turn, has potentially important implications for the direction of technological change, since inventors just like Cochran, are disposed towards solving problems that they actually face.

To shed light on this question of who invents, we build a dataset of inventors matching individuals granted patents between 1870 and 1940 to the corresponding decennial US Population Census. Our contribution in terms of data construction is two-fold. First, we construct a new dataset of individual patents and patentees¹ from the *Annual Reports of the Commissioner of Patents*. Our second data contribution is a set of patentees matched to the Census, which provides a variety of demographic and economic information on patentees such as age, race, gender, marital status, migrant status, and occupation.

With this dataset, we document time series patterns on the demographics of inventors.² Unsurprisingly, inventors on the whole are not a random sample of individuals. They also differ in some less obvious ways such as being older and more likely to have been born outside the U.S. When we look over time, a number of these patterns are remarkably stable. For example the odds, relative to population representation, that an inventor is not white

¹We use the term “patentee” and “inventor” interchangeably.

²In the appendix, we study the cross-sectional determinants of the demographic of inventors in a regression framework.

is around 0.29 in 1880 and only 0.34 in 1940. The relative age of patentees is also remarkably persistent over the 70-year period we study, with patentees being 11-14% older than the average population of their counties. Temin (1999) finds a similar persistence in the demographics of the American business elite. On the other hand, the relative odds that an inventor is female, while always modest, fluctuates considerably and almost doubles over this period, ranging from a low of 0.07 in 1880 to a high of 0.13 in 1940. To be clear, this says that women, who comprise about 50 percent of the population in all periods, make up between 3.5 and 6.5 percent of inventors. These demographic patterns (persistence in the case of non-whites and slow growth in the case of women) takes place against a backdrop in which the skills exhibited by inventors is changing rapidly: we document a rapid fall in the relative odds that an inventor is working in farming and a rise in the relative odds that a patentee works in a white-collar occupation.

We also examine a number of characteristics related to the migration status of a patentee. Over the whole period, they are relatively more likely to: (1) be living outside their state of birth; (2) to have been born outside the U.S. and (3) have parents who were born outside the U.S. What is perhaps more interesting is that this overrepresentation increases over time. For example in 1870, it is only 1.01 times more likely that an inventor will be foreign born, and, by 1940, this ratio is around 1.65.

When we extrapolate the trends in demographics from the late 19th and early 20th century to the present day, we find predicted rates of patenting for women and non-whites remarkably similar to the actual rates today (Jones, 2009; Cook and Kongcharoen, 2010; Bell et al., 2018). This is particularly startling given that since the end of our sample in 1940, there have been major institutional changes that increased access to civil rights, education, and employment among women and racial minorities in the United States. Yet this formal access has apparently not translated to an acceleration of the rate of participation of these groups in patenting.

We are not the first to study the demographics of inventors in the American past

(Schmookler, 1957; Sokoloff, 1988; Sokoloff and Khan, 1990; Lamoreaux and Sokoloff, 2005). What distinguishes our work is the scale of our analysis. Our dataset is novel in its size and scope, in terms of time period covered and information on patentees.³ As one point of comparison, work by Khan and Sokoloff (1993) and Cook (2011) has attempted to document inventor demographics by consulting detailed biographies of individual inventors. These biographies can provide more detail than what the Census provides, but this process to learn inventors’ identities is time consuming and difficult to scale. It also raises concerns that particular groups may be systematically omitted from these sources. While our approach is not without drawbacks, we are able to overcome the challenges of the biography approach by starting with the universe of patentees rather than a selected sample for which a biography was written. This ensures a final sample of inventors that is more reflective of the whole population of inventors.

2 Data Construction

2.1 Generating a List of Patent Grantees

Our list of patent grantees comes from the *Annual Report of the Commissioner of Patents*. This was an official document published annually that includes among other things every patent granted in a given year and the individuals to whom it was granted. We first convert images into plaintext documents using optical character recognition (OCR) software. We then extract from these plaintext documents the relevant information on names and locations of inventors. To do this, we use an algorithm comprising 35 regular expressions to handle the myriad ways that the information for a particular patent can be presented.⁴ We score the

³In contemporaneous work, Akcigit et al. (2017) assemble a dataset similar to ours. They use this data to study the relationship between patented inventions, and long-run growth and inequality, whereas our paper is focused on understanding the evolution of inventor demographics and its implications for the trajectory of innovation.

⁴We are happy to provide the Python code used in this parsing process. Additional details of the parsing process are presented in the Appendix.

results from these different regular expressions in deciding which results to accept. Because our final aim is to match these patentees to the Census, we place a high value on regular expressions that return a valid state and non-empty values for first and last names.⁵ Note that our parsing strategy attempts to identify all individuals on a particular patent, not just the first one listed.

Figure 1 shows the rate at which we extract potentially matchable inventors as a percentage of the total number of patents reported by the U.S. Patent and Trademark Office (USPTO) for the Population Census years between 1870 and 1940. There are, unsurprisingly, fluctuations over time in this rate though no clear trend is visible. These could be due to variation over time in the quality of the plaintext and in systematic ways in which the data are formatted in the *Annual Reports*. The images of the original reports do seem to have a particularly high number of unreadable pages in 1910 and 1930, the two years with the lowest parse rate. Nevertheless, the parse rate is high across all years, ranging from a low of just under 70% of all patents in 1930 to a high of about 90% in 1880.

In the appendix, we discuss in greater detail the problems in going from the initial plaintext of the *Annual Reports* to this parsed list of patent grantees. We provide a number of checks to show that we are not systematically excluding particular types of inventors due to our parser or the matching procedure. To show that the parsing does not introduce systematic biases, we compare our list to a “definitive” list collected by Dr. Jim Shaw of Hutchinson, KS from the Subject-Matter Index⁶ overlapping our dataset over the first 5 years. We also discuss other potential data sources for patent records comparing and contrasting those to ours.

⁵As we discuss below, we also use town of residence in matching the inventors. Because of the difficulty in deciding what is a “valid” town name (versus one of only 51 state names including Washington, D.C.), we do not place any weight on this part of the output when scoring each regular expression.

⁶Dr. Shaw hand-transcribed all patents granted for the years 1790 to 1873.

2.2 Matching Patentees to the Census

We match the list of “parsed” patentees to the Census of Population in 1870, 1880, 1900, 1910, 1920, 1930, and 1940.⁷ The decennial Census offers demographic information such as race, age, gender, and birthplace for the individual. We “fuzzy” match on the first and last name of the patentee as well as the town listed on the patent records. We require an exact match on the state of the residence as well as first letter of first and last names. The idea behind our matching strategy is that the town listed on the patent records is (most likely) the town where the inventor is living at the time of the census. On this geographic dimension, we have more information than other matching exercises such as Long and Ferrie (2013) that search for individuals across the whole country. On the other hand, the information we are matching on is rather limited relative to other work in the literature that also uses information on age and birthplace (Long and Ferrie, 2013; Abramitzky et al., 2012). To select weights for the variables used in the fuzzy matching procedure, we manually matched patentees to the census for a subset of the data to attempt to identify the “ground truth” for this subsample (Bailey et al., 2019).⁸ We then chose weights to give an appropriate balance between false positive and false negative matches.⁹

There are two things to note about our matching procedure. First of all, we do not require the matching to be “injective.” This means that the same person in the Census may be a potential match on multiple patents. Given that someone could receive multiple patents in a single year, it makes sense to allow for this possibility. Second, our matching procedure returns all “plausible” matches and we will not enforce unique matching. There are several ways to handle the resulting cases of multiple potential matches. First, we construct “best”- and “worst”-case bounds on statistics of interest using the data on all potential matches. For example, we calculate the upper and lower bounds on the average age of inventors by taking

⁷The 1890 population census manuscripts were destroyed in a fire, and so we cannot match from the patent records to this census.

⁸In particular, we used the population of 1900 Vermont and searched for all patentees by hand.

⁹We also excluded all individuals younger than 15 or older than 80 under the assumption that very few of these individuals obtain a patent in any given year.

the average of the maximum and minimum ages, respectively, over all possible matches by patentee. We also average over all possible matches as suggested by Poirier and Ziebarth (2018). This treats all the possible matches as “exchangeable” and hinges on the fact that with probability one, the correct match is in the set of possible matches.¹⁰

Figure 2 plots some statistics of our matching procedure over time. In calculating the percentages, the denominator is the number of “parsed” inventors, which is the numerator of the percentages reported in Figure 1, rather than the number of patents or the number of patents with at least one parsed inventor. We also report the percentage of inventors that are uniquely matched to an individual in the Census and the number of “perfect” matches, which are matches in which all of the characters in the last name, first name, and town of an inventor match exactly with an individual in the census. These two sets need not be strict subsets of one another.

Over the whole period, we match around 10% of all patentees with the vast majority of those being unique and about 70% of them being perfect. The average number of matches increases from around 2.5 individuals in the 1870 Census to 4 individuals in 1940. In thinking about these rates, it is useful to compare and contrast our setting with other leading work linking individuals over time such as the papers by Long and Ferrie (2013) and Abramitzky et al. (2012). The focus of these papers is different (intergenerational mobility) which leads them to match individuals in consecutive Population Censuses. This is quite different from our case where we match from the patent data in a given year to the Census in that year. This provides them with additional identifying information such as age and place of birth while at the same time requiring them to search the whole population to allow for individuals’ geographic mobility. These other papers usually report match rates (where a match is also required to be unique) of around 30%. Saying whether our rate is “too” high or low is challenging given the differences in variables we use for matching as well as the differences

¹⁰An additional approach is to select the “best” match, which is the match with the highest match score; when multiple individuals produce the same match score, we randomly pick one as the best match. This produces results nearly identical to those when we take the average over all possible matches, and so we do not plot them below.

in the quality of the records attempting to be linked.

To allay some concerns about the matching process, in the appendix we compare observable characteristics including first and last name of matched patentees to those unmatched and find that these distributions look similar. We also offer a robustness check varying how high the match score must be for us to declare an individual as a potential match. Finally, we address the problem that there is a lag between when a person files a patent and when the patent is granted. The location listed on the *Annual Reports* is that of filing, but if a person moved between when the patent was filed and when it was granted, then any matches we find for this person would be in error. To examine the extent of this problem, we match years of the *Annual Reports* that follow a given Population Census. For example, we match the 1881 *Annual Reports* to the 1880 Population Census. Matching to adjacent years does not substantially affect the results.

3 Changes in the Demographics of Patentees Over Time

We now document time-series patterns in the demographics of patentees over this time period. We report the odds ratio of a particular demographic characteristic relative to the US population conditional on individuals being between 15 and 80 years of age. Recall that we impose this age restriction to begin with when searching for patentees in the Census records. A value of this ratio greater than 1 means a particular demographic group is overrepresented among inventors than would be expected by chance. Figure 3 shows that inventors are demographically quite different from the general population. Some of these differences are not surprising such as the fact that non-whites and women are underrepresented. The fact that these results fit our priors on who patents provides a quality check on our matching procedure. If it had been the case that the demographics of inventors was similar to the demographics of the population overall, this would have been suggested that we were simply randomly matching inventors to individuals in the population.

Panel (a) shows little change in the relative average age of patentees in our sample. Over the seventy-year time frame, patentees are on average 41 years old as compared to the population average of 37. This stability is in contrast to recent work documenting that the age of innovators including but not limited to patentees has been steadily increasing over the last 40 years. For example, the age of first time NIH grant recipients has steadily increased from 37 in 1980 to 42 in 2008 (Kaiser, 2008). Jones (2009, 2010) show that the age of first invention has been increasing between 1985 and 2000. He attributes this trend, at least in part, to the increased time necessary to acquire the human capital used in the invention process. However, there is some evidence that since 2000, the average age of all inventors and first-time inventors in Sweden has been declining quite rapidly from a peak of just over 46 in 1997 to 43 in 2007 (Jung and Ejermo, 2014). This trend raises the possibility that the patterns observed by Jones between the mid 1980s to 2000 were simply transitory and that the ages of inventors are returning to something like they were in 1900, when patentees were just over 10% older than the average age in the US population. In all these other studies, the results on age are difficult to directly compare to ours, either because they examine (1) age at first invention (Jones, 2009); (2) a set of highly selected inventors (Jones, 2010); or (3) invention in another country (Jung and Ejermo, 2014). Bell et al. (2018) presents results from patentees in recent decades that are more comparable to our sample of historical patentees. They find that the average age of inventors is 43.7 from 1996 to 2014, almost identical to the 43.5 years in our sample in 1940.¹¹ Moreover, the persistence of inventors being in their forties is consistent with recent work by Galenson (2016); Azoulay et al. (2018).

Next, panel (b) shows that 97% of patentees are white versus around 90% of the population while panel (c) shows 95% of patentees are male as compared to 51% in the general population. The term non-white encompasses not only African Americans but any other racial minority. Recent evidence suggests the demographics of inventors in the 21st century

¹¹One minor difference is that the age used in Bell et al. (2018) is the patentee age at the time of patent filing, whereas ours is age at which the patent was granted. Since the average time between patent filing and issuance was typically short over the period we study (Berkes, 2018), these ages should be similar.

do not look dramatically different from the late 19th and early 20th centuries. Women and blacks to this day account for a disproportionately low fraction of inventive activity. Bell et al. (2018) find that women account for about 13% of patent grantees between 1996 and 2014.¹² Using different methods to determine the gender of patentees, Milli et al. (2016) and Toole et al. (2019) find nearly identical shares of female inventors. Ding et al. (2006) find that only 5.65% of female scientists patent at all versus 13% for males, and females hold only about 6% of filed patents in their sample. Moreover, females are currently both less likely to obtain patent rights after filing patent applications and to maintain those rights once they obtain them (Jensen et al., 2018). While some studies such as Ding et al. (2006) and Frietsch et al. (2009) find declining gender gaps, this catch-up is slower than female engagement in other comparable parts of society such as PhD education in Science and Engineering (Jung and Ejermo, 2014) and even in high skilled occupations like doctors and lawyers (Hsieh et al., 2019). Similarly, the representation of blacks in patenting activity today remains dismal (Cook and Kongcharoen, 2010; Bell et al., 2018).¹³

On the other hand, the representation of these groups is not static over our time period. While the overall share remains low, women just about doubled their representation in patenting between 1870 and 1940. While such growth is encouraging, if we project the 1870-1940 growth rate in female patenting forward in time, we would expect females to account for about 11% of patents in 2010, which is very similar to the 13% Bell et al. (2018) find in their study. Thus, little has occurred since 1940 to accelerate the rate at which the gender gap in patenting is closing. Non-whites also increased their representation during the period we study, although only slightly.¹⁴ Particularly surprising for this group is the fact that there is no discernible change with the end of Reconstruction or the beginning of the Jim Crow

¹²This number is taken from Table 1 of Bell et al. (2018) as are the other numbers cited from this paper.

¹³Panel (d) shows that 72% are married versus 60% of the population. We are not aware of any literature on this particular demographic.

¹⁴One drawback of matching to the Population Census is that we do not use the *Annual Reports* in the years between decennial censuses. As a way to get a more complete time series picture, in the appendix we offer a second method akin to two-sample IV that uses each patentee’s first name to infer their race and gender.

era, events that saw tremendous setbacks for African-Americans along economic, social, and political dimensions.¹⁵ When we project the growth of non-white patenting forward in time, we would predict an odds ratio of about 40% in 2010, while Bell et al. (2018) find an odds ratio of 67%, suggesting that the racial patenting gap has closed faster between 1940 and today than it did between 1870 and 1940.¹⁶ But even here, the case for optimism is limited: the composition of racial groups in the U.S. population has been changing over time, with individuals identified as Asian accounting for a growing share of the population and also patenting at disproportionately high rates. When Asians are excluded from the paper by Bell et al. (2018), the odds ratio for blacks and Hispanics is 37%, less than we would expect if the trends from 1870 to 1940 continued and in fact strikingly similar to the 34% we observe in 1940.

Panels (e) and (f) of Figure 3 examine the limited information we have on occupations of patentees.¹⁷ The first of these is whether a person works on a farm as defined by whether the term “farm” is used in the occupation or industry string. The white-collar occupation variable is whether a person works in this type of job, which is defined based on the 1950 occupational codes and is only available after 1900. We find a shift away from working on a farm to being in a white-collar occupation. White-collar occupations actually go from being underrepresented in 1900 to overrepresented relative to the population at large by 1940. These shifts are over and above the shifts in the general population away from agriculture and towards white-collar work. We think these results reflect not only changes in the types of innovations being made and the propensity to patent, but also the skills necessary to make those innovations. This interpretation is consistent with the finding in the paper by Bloom et al. (2017) that the skill and research effort requirements for invention have increased

¹⁵In the appendix, we plot odds ratios for non-white patentees separately for northern and southern states. The dynamics of non-white patenting is similar in both regions though the overall level is much lower in the south.

¹⁶The race data in Bell et al. (2018) come from the sample of patentees who attended New York Public Schools and thus may not be representative of the racial characteristics of patentees in the entire country.

¹⁷While there is an occupation variable in the 1880 Population Census, it is recorded as a raw string making it very difficult to classify occupations into white- versus blue- collar. So we do not report any results for this variable in that year. There is no occupation variable in 1870.

since the 1930s, and our findings suggest that these trends might have even begun in the decades before. The changes in the occupations of inventors, and the corresponding skills that inventors possess, draws a striking contrast to the persistence in patenting for non-whites and, to a lesser extent, women, particularly given the changes in human capital and labor force participation for these groups over the period we study.

Even more striking are the changes reflected in Figure 4 which examines a number of migration related variables. We define an “interstate migrant” as a person living in a state other than where he was born. Unfortunately, we cannot identify how long a particular inventor has been living in a particular location before patenting or any finer geographic detail on where the person moved from since for people born in the US all that is reported in the state of their birth. We also examine whether an inventor was foreign born as well as for the inventor’s mother and father.¹⁸

Inventors in this period were more “mobile” along all four of the migration measures we consider. In 1870, the relative odds of being an interstate migrant are only slight greater than one, but, by 1940, a patentee is around 1.6 times more likely to be an interstate migrant. Similarly, in 1940, 21.5% of patentees were foreign born relative to 13% of the population, which was a major increase in relative terms from 1870 when patentees were foreign born at about the same rate as the population as a whole. The same qualitative patterns for the nativity of an inventors’ parents are also evident.

Our data cover an eventful period in the history of international immigration to the US, including most of the Age of Mass Migration and the strict immigration restrictions imposed in 1917 and 1921. In spite of this sharp change in immigration policy, we find no corresponding change in the trend of patenting for foreign born people following the imposition of quotas on certain nationalities. Existing research on this wave of immigrations (Abramitzky et al., 2012, 2013; Abramitzky and Boustan, 2017) finds that they were negatively selected on average, at least, in terms of income. While understanding the averages

¹⁸Unfortunately, we do not have the place of birth of a person’s parents in 1870.

of the immigrant population is important, inventors tend to come from the right tail of the ability distribution (Bell et al., 2018; Aghion et al., 2018). Our results show that, in terms of inventiveness, the right tail of the distribution for migrants was thicker than that for natives. This is consistent with contemporary patterns that highlight the large contribution of immigrants to innovation in 20th century America (Wadhwa et al., 2007; Kerr and Lincoln, 2010; Hunt and Gauthier-Loiselle, 2010; Kerr, 2013, 2018).

In the appendix, we study the cross-sectional patterns of demographics of inventors. In particular, we regress the demographics of patentees on county-level predictors including representation of that particular demographic group and other economic and demographic characteristics. The only consistent predictor of the representation of a particular demographic is the representation of that group in the local population overall.

4 Conclusion

Our empirical findings on the demographics of patentees over this period present a complex picture of stability along some dimensions such as age, change along others such as migration, and continued underrepresentation for others such as women and non-whites. In fact, these gender and race gaps appear to be closing no faster today than they did between 1870 and 1940. We leave many open questions for future work. Why did immigrants become better poised to participate in invention than women and blacks? What is the relationship between institutional change in the form of civil rights and political representation, and participation in invention by the underrepresented? How important is human capital in the form of formal education for invention?

Finally, how important is who invents for understanding the process of technological change itself? New inventions primarily build off prior ones generating path dependence, and inventors necessarily seek to solve problems with which they are familiar. Therefore, the lack of demographic or socio-economic inclusiveness in the process of invention affects

not just the rate of innovation, but presumably also the direction and scope of technological progress in the form of what problems are considered important. It is our hope that the data introduced in this paper can be used to shed light on these questions.

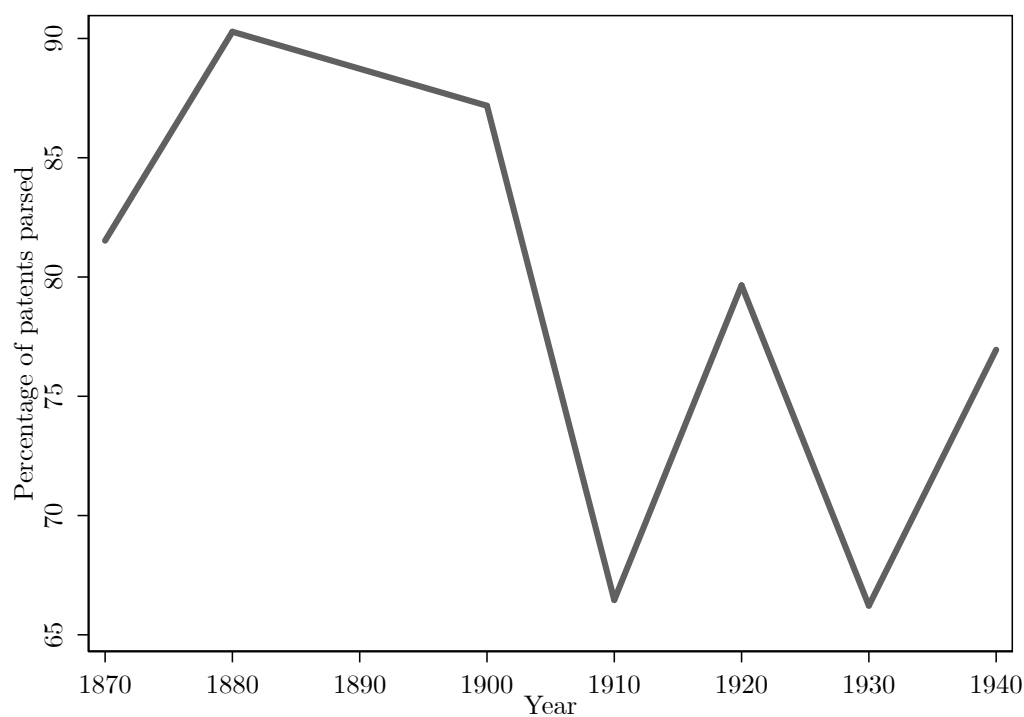
References

- Abramitzky, R. and L. Boustan (2017). Immigration in American economic history. *Journal of Economic Literature* 55, 1311–1345.
- Abramitzky, R., L. P. Boustan, and K. Eriksson (2012). Europe’s tired, poor, huddled masses: Self-selection and economic outcomes in the Age of Mass Migration. *American Economic Review* 102, 1832–1856.
- Abramitzky, R., L. P. Boustan, and K. Eriksson (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the Age of Mass Migration. *Journal of Development Economics* 102, 2–14.
- Aghion, P., U. Akcigit, A. Hyytinen, and O. Toivanen (2018). The social origins and IQ of inventors. Unpublished, Harvard University.
- Akcigit, U., J. Grigsby, and T. Nicholas (2017). The rise of American ingenuity: Innovation and inventors of the Golden Age. NBER Working Paper 23047.
- Azoulay, P., B. Jones, J. D. Kim, and J. Miranda (2018). Age and high-growth entrepreneurship. NBER Working Paper 24489.
- Bailey, M., C. Cole, M. Henderson, and C. Massey (2019). How well do automated linking methods perform? Lessons from U.S. historical data. Unpublished, University of Michigan.
- Bell, A. M., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2018). Who becomes an inventor in America? The importance of exposure to innovation. *Quarterly Journal of Economics* Forthcoming.
- Berkes, E. (2018). Comprehensive Universe of U.S. Patents (CUSP): Data and facts. Unpublished, Ohio State University.
- Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2017). Are ideas getting harder to find? NBER Working Paper 23782.
- Cook, L. D. (2011). Inventing social capital: Evidence from African American inventors, 1843–1930. *Explorations in Economic History* 48, 507–518.
- Cook, L. D. and C. Kongcharoen (2010). The idea gap in pink and black. NBER Working Paper 16331.
- Ding, W. W., F. Murray, and T. E. Stuart (2006). Gender differences in patenting in the academic life sciences. *Science* 313, 665–666.
- Frietsch, R., I. Haller, M. Funken-Vrohlings, and H. Grupp (2009). Gender-specific patterns in patenting and publishing. *Research Policy* 38, 590–599.
- Galenson, D. (2016). Creative life cycles: Three myths. Becker Friedman Institute for Research in Economics Working Paper 2016-28.

- Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and US economic growth. *Econometrica Forthcoming*.
- Hunt, J. and M. Gauthier-Loiselle (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics* 2, 31–56.
- Jensen, K., B. Kovács, and O. Sorenson (2018). Gender differences in obtaining and maintaining patent rights. *Nature Biotechnology* 36, 307–309.
- Jones, B. F. (2009). The burden of knowledge and the “Death of the Renaissance Man”: Is innovation getting harder? *Review of Economic Studies* 76, 283–317.
- Jones, B. F. (2010). Age and great invention. *Review of Economics and Statistics* 92, 1–14.
- Jung, T. and O. Ejermo (2014). Demographic patterns and trends in patenting: Gender, age, and education of inventors. *Technological Forecasting and Social Change* 86, 110–124.
- Kaiser, J. (2008). *Science* 322, 834–835.
- Kerr, W. and W. Lincoln (2010). The supply side of innovation: H-1B visa reforms and US ethnic invention. *Journal of Labor Economics* 28, 473–508.
- Kerr, W. R. (2013). U.S. high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence. NBER Working Paper 19377.
- Kerr, W. R. (2018). *The Gift of Global Talent: How Migration Shapes Business, Economy & Society*. Stanford Business Books.
- Khan, B. Z. and K. L. Sokoloff (1993). “Schemes of Practical Utility”: Entrepreneurship and innovation among “Great Inventors” in the United States, 1790-1865. *Journal of Economic History* 53, 289–307.
- Lamoreaux, N. R. and K. L. Sokoloff (2005). The decline of the independent inventor: A Schumpeterian story. NBER Working Paper 11654.
- Long, J. and J. Ferrie (2013). Intergenerational occupational mobility in Great Britain and the United States since 1850. *American Economic Review* 103, 1109–1137.
- Milli, J., E. Williams-Baron, M. Berlan, J. Xia, and B. Gault (2016). Equity in innovation: women inventors and patents. Technical report, Institute for Women’s Policy Research.
- Poirier, A. and N. L. Ziebarth (2018). Estimation of models with multiple-valued explanatory variables. *Journal of Business and Economic Statistics Forthcoming*.
- Schmookler, J. (1957). Inventors past and present. *Review of Economics and Statistics* 39, 321–333.
- Sokoloff, K. L. (1988). Inventive activity in early industrial America: Evidence from patent records, 1790-1846. *Journal of Economic History* 48, 813–850.

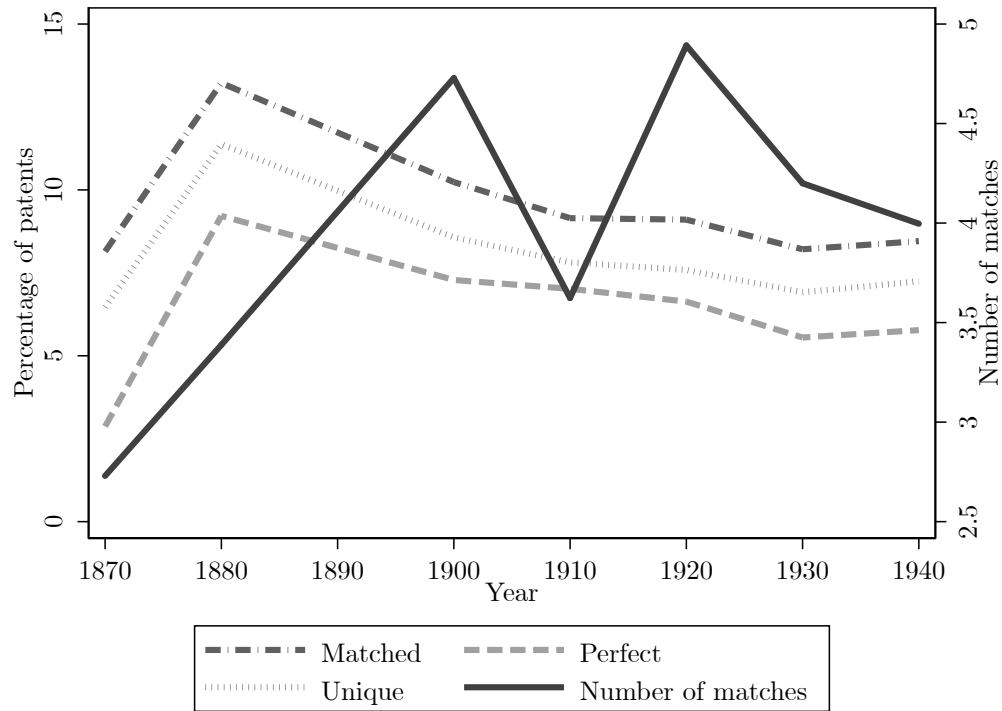
- Sokoloff, K. L. and B. Z. Khan (1990). The democratization of invention during early industrialization: Evidence from the United States, 1790-1846. *Journal of Economic History* 50, 363–378.
- Temin, P. (1999). The stability of the American business elite. *Industrial and Corporate Change* 8, 189–209.
- Toole, A. A., S. Breschi, E. Miguelez, A. Myers, E. Ferrucci, V. Sterzi, C. A. W. deGrazia, F. Lissoni, and G. Tarasconi (2019). Progress and potential: a profile of women inventors on U.S. patents. Technical report, U.S. Patent & Trademark Office, Office of the Chief Economist.
- Wadhwa, V., A. Saxenian, B. Rissing, and G. Gereffi (2007). America’s new immigrant entrepreneurs. Technical report, University of California Berkeley.

Figure 1: Parsing Rates of *Annual Reports* in Census Years



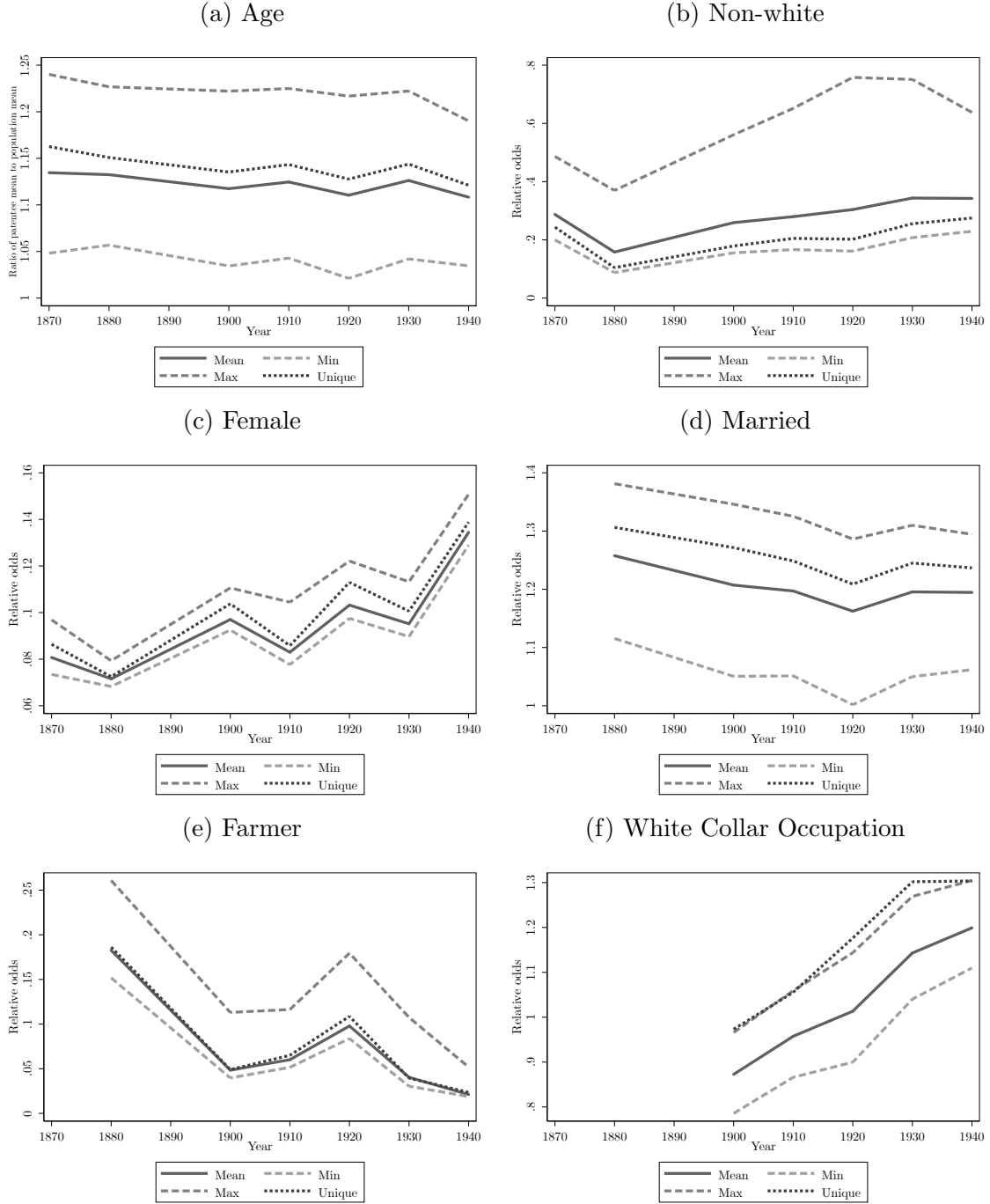
Notes: The percentage is calculated using the number of patents granted in the given year as reported by *Annual Report of the Patent Commissioner* as the denominator. A patent is “parsed” if our parser returns a non-blank first and last name as well as a valid state name for at least one patentee.

Figure 2: Matching Statistics in Census Years



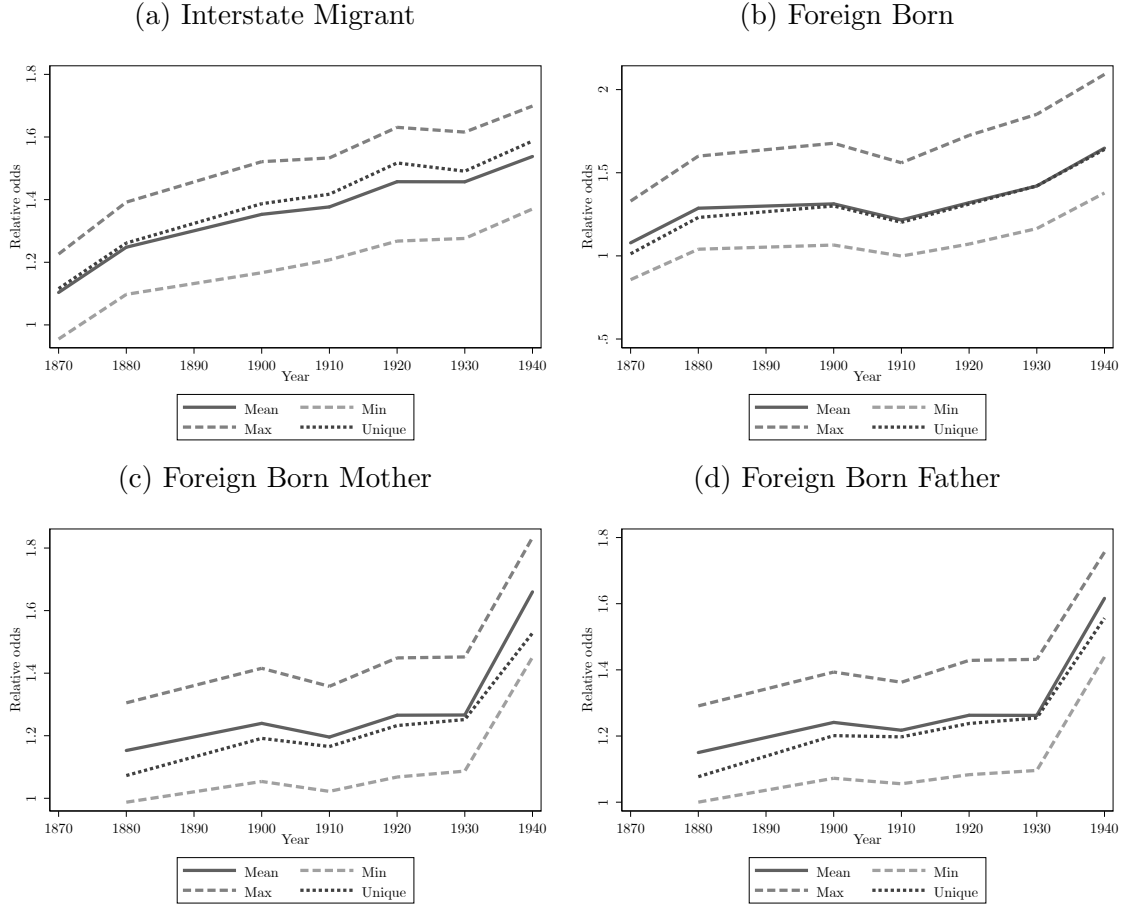
Notes: All percentages are as a fraction of the total number of patentees that were “parsed” in a given year. “Parsed” means that a patentee has a non-blank first and last name string as well as a valid state name. This set is the numerator of the percentages reported in Figure 1. “Matched” are those for which we can find at least one match. “Perfect” are those patentees for which we can find an exact match to a person’s name and town of residence. Note here as well that this need not be a unique match. “Unique” are those patentees that have a unique match meaning there is no other person in the town of the patentee with a name sufficiently close. Note that this match need not be perfect. The average number of matches (the right axis) is the average number of individuals in the Census that are matches for a given patentee.

Figure 3: Demographics of Patentees Relative to Population



Notes: The “Mean” line is the (equally weighted) average characteristic across all possible matches for a given patentee. The “Max” line (resp. “Min”) is the maximum (resp. “minimum”) value of a particular characteristic across all possible matches for a given patentee. We then report the average of the maxima (resp. minima) over all patentees. We also report the averages for the set of unique matches (the “Unique” line). For each of these, we report the ratio relative to the average demographic characteristic in US in the respective year.

Figure 4: Demographics of Patentees: Migration



Notes: The “Mean” line is the (equally weighted) average characteristic across all possible matches for a given patentee. The “Max” line (resp. “Min”) is the maximum (resp. “minimum”) value of a particular characteristic across all possible matches for a given patentee. We then report the average of the maxima (resp. minima) over all patentees. We also report the averages for the set of unique matches (the “Unique” line). For each of these, we report the ratio relative to the average demographic characteristic in US in the respective year.