The Contribution of High-Skilled Immigrants to Innovation in the United States

Shai Bernstein, Rebecca Diamond, Timothy McQuade and Beatriz Pousada*

July 11, 2019

Abstract

We characterize the contribution of immigrants to US innovation, both through their direct productivity as well as through their indirect spillover effects on their native collaborators. To do so, we link patent records to a database containing the first five digits of 236 million of Social Security Numbers (SSN). By combining this part of the SSN together with year of birth, we identify whether individuals are immigrants based on the age at which their Social Security Number is assigned. We find that over the course of their careers, immigrants are more productive than natives, as measured by number of patents, patent citations, and the economic value of these patents. Immigrant inventors are more likely to rely on foreign technologies, to collaborate with foreign inventors, and to be cited in foreign markets, thus contributing to the importation and diffusion of ideas across borders. Using an identification strategy that exploits premature inventor deaths, we find that immigrants collaborators create especially strong positive externalities on the innovation production of their collaborators, while natives have a much weaker impact. We identify a key mechanism driving these differences: unique knowledge backgrounds of immigrants. This suggests that combining different knowledge pools into inventor teams is important for innovation.

*Shai Bernstein is with Stanford University GSB, and NBER, Rebecca Diamond is with Stanford University GSB, and NBER, Timothy McQuade is with Stanford University GSB, and Beatriz Pousada is with Stanford University. The authors have obtained IRB approval from Stanford University before conducting the analysis.

1 Introduction

Innovation and technological progress is thought to be a key determinant of economic growth (Aghion and Howitt (1992); Romer (1990)). There is growing suggestive evidence that immigrants play a key role in US innovation. For example, immigrants comprised 23% of the total workforce in STEM occupations in 2016.[1] They account for 26% of US-based Nobel Prize winners from 1990 through 2000. Based on a 2003 survey, US immigrants with a 4-year college degree were twice as likely to have a patent than US-born college grads (Hunt and Gauthier-Loiselle (2010)).

Despite this suggestive evidence, we do not have a comprehensive estimate of how much immigrants contribute to US innovation, as measured by patents. In this paper, we bring to bear new data and utilize a unique approach to identify the immigrant status of individuals residing in the United States, which we then link to patent data. We find immigrants account for 16% of all US inventors from 1976 through 2012. Immigrants, however, account for about 23% of total innovation, as we find the average immigrant is substantially more productive than the average US-born inventor.

These metrics account for the direct output differences of immigrant and native inventors. We further investigate whether immigrants create spillovers onto the innovation of native inventors, thus indirectly contributing to innovation by raising native inventor productivity. To investigate this mechanism, we use unexpected early deaths of native and immigrant inventors as a source of causal variation in number of native/immigrant collaborators other inventors have access to. We find collaborating with immigrants lead to especially large productivity gains for inventors, relative to collaborating with US natives. Our results suggest that immigrants and natives draw on different knowledge pools, the combination and sharing of which is especially fruitful for innovation.

Our analysis relies on the Infutor database, which provides the exact address history of more than 236 million adults living in the United States over the past 30 years. Beyond the exact address history, this data also includes the individuals' names, years of birth, genders, and the first 5-digits of their Social Security numbers. Our methodology infers immigrant status by combining the first 5-digits of the SSN together with information on year of birth. The first five digits of the SSN pin down the year in which the SSN was assigned. Since practically all US natives are assigned a SSN during their youth, or even at birth, those individuals who receive a SSN in their twenties or later are highly likely to be immigrants. We validate our method with data from the Census and American Community Survey (ACS) and find we capture the cross-sectional variation in immigrant shares across US counties, with R^2 of around 90% across multiple specifications.[2]

Using individual-level address information provided by both Infutor and the USPTO, we merge information on an individual's immigrant status with the universe of patents. We find that 16% of all US-based inventors, between 1976-2012, have been immigrants that came to the United States when

[1] Data are from the 2016 American Community Survey. STEM occupation defined as engineers, mathematical and computer scientists, natural scientists, and physicians.

[2] Our method only identifies immigrants that have legal status. Since our interest is in studying the innovative contributions of high-skilled immigrants working in US companies, this is not a significant limitation.

they were 20 years of age or older. The contribution of these immigrants to overall US innovative output, however, have been disproportionate relative to their share of the US inventor population. Immigrant inventors have produced roughly 23% of all patents during this time period, more than a 40% increase relative to their share of the US-based inventor population. These patents, moreover, do not appear to be of lower quality. Using the number of patents weighted by the number of forward citations, which captures the quality of innovation (Hall et al. (2001)), we find that the immigrant contribution is even higher at 24%. Finally, using the Kogan et al. (2017) measure capturing stock market reaction to patent grants, we find that the immigrants have generated 25% of the aggregate economic value created by patents produced by publicly traded companies, an increase of 47% relative to their share of the inventor population working in publicly traded companies.

The contribution of immigrants to US innovative output is not particularly concentrated in specific sectors. We find that immigrants generate about 25% of innovative output in the Computers and Communications, Drugs and Medical, Electronics, and Chemical sectors, but only 15% in more traditional technologies such as metal working, transportation, and engines.

We next explore how immigrants differ in their innovative productivity over the life-cycle. Both natives and immigrants exhibit an inverse U-shape pattern, where inventors are quite unproductive at the beginning of their careers, become most productive in their late 30s and early 40s, and then steadily decline in productivity thereafter.[3] However, while the two populations follow similar trajectories, immigrants diverge from natives when reaching to the peak of innovative productivity, with immigrants producing significantly more patents, citations, and generating more economic value. This gap persists throughout the rest of their careers. These differences are also quite similar across cohorts of inventors.

While the goal of this paper is not to fully decompose all the reasons immigrants are more productive than natives, we do investigate a few mechanisms. While immigrant inventors in the US may simply be selected based on their innate ability, we do observe them also making choices that complement their productivity. For example, we find immigrants are disproportionately choosing to live in highly productive counties ("innovation hubs"), relative to US born inventors. Immigrants also are disproportionately patenting in technology classes that are experiencing more innovation activity. These two forces can explain about 30% of the raw patenting gap between immigrants and natives. This suggests that immigrants are not only more productive based on ability, but that they are more willing make choices that further improve their innovative output.

We find that immigrant inventors foster the importation of foreign ideas and technologies into the United States and facilitate the diffusion of global knowledge. During their careers, immigrant inventors rely more heavily on foreign technologies, as illustrated by a ten percent increase in the fraction of backward foreign citations. Immigrants are also about twice as likely to collaborate with foreign inventors, relative to native inventors. Finally, foreign technologies are about ten percent

[3] These findings hold with respect to patent production, the citation adjusted number of patents, and the economic value of the patents produced. These inverse U-shape productivity patterns are consistent with a large literature exploring the relationship between age and scientific contributions (see Jones et al. (2014a) for a survey), reflecting the necessary time to accumulate relevant human capital

more likely to cite the patents of US-based immigrants relative to US natives.

While US-based immigrant inventors appear to be more productive than US natives, one potential concern is that, due to cultural impediments or lack of assimilation, immigrant inventors may be less integrated into the overall US knowledge market, may remain isolated at their workplace, and thus may contribute less to the team-specific capital which Jaravel et al. (2018) document is important to the innovative process. In contrast, we find that throughout their career, immigrant inventors tend to have more collaborators than native inventors. Furthermore, while we do find that immigrants are more likely to work with other immigrants (as compared to natives), this tendency declines over the life-cycle, suggesting a gradual assimilation process.

These team interactions between foreign and US born inventors in the production of patents are of particular interest since they may be a key mechanism through which an inventor's knowledge spills over onto the knowledge and productivity of his collaborators. These knowledge externalities are exactly why the US may be able to allow high-skilled immigrants in the country and improve the welfare and productivity of US-born workers. We estimate the magnitudes of foreign born and US born knowledge externalities on their collaborators using the exogenous termination of such relationships. Specifically, to construct causal estimates of these spillovers, we exploit the premature deaths of inventors, defined as deaths that occur before the age of 60.[4] We then follow the patenting behavior of inventors who had co-authored a patent with the deceased inventor, at some point prior to the inventor's death. We compare the change in patenting activity of these co-authors before versus after the inventor death to a matched control group of inventors who did not experience the pre-matur death of a co-author. This form of identification strategy is becoming increasingly common in the literature (Jones and Olken, 2005; Bennedsen et al., 2008; Azoulay et al., 2010; Nguyen and Nielsen, 2010; Oettl, 2012; Becker and Hvide, 2013; Fadlon and Nielsen, 2015; Isen, 2013; Jaravel et al., 2018).

Overall, we find that premature death leads to a 12 to 15 percent decline in the innovative productivity of their co-inventors, as measured by patents and top patents, consistent with Jaravel et al. (2018). This decline takes place gradually, and has a long lasting impact. Most strikingly, we find that the disruption caused by an immigrant death causes a significantly larger decline in the productivity of the co-inventors than that of native inventors. The death of an immigrant lowers co-inventor productivity by approximately 26%, while a US-born inventor death lowers productivity by approximately 10%. These gaps are large, persistent, and take place across all of our measures of innovative productivity.

Our paper contributes to several strands of literature. It is most directly linked with a growing literature that evaluates the effects of high-skilled immigration on innovation. This literature has been constrained by the limited availability of individual-level data on the immigrant status of innovative workers. A few papers have relied on ethnic-name databases to classify scientists with names associated with specific foreign countries as immigrants (e.g., Kerr (2010); Kerr and Lincoln

[4]We link our data to a public-use copy of the social security death master file to identify inventor deaths courtesy of SSDMF.INFO.

(2010); Kerr (2008a,b); Foley and Kerr (2013)). However, as pointed out by Kerr (2008b), this method introduces significant measurement error and cannot differentiate foreign-born individuals from US natives with ethnic names. It also cannot identify immigrants from Western Europe. A few papers have used the 2003 National Survey of College Graduates to measure patenting differences between immigrants and natives (Hunt and Gauthier-Loiselle (2010), Hunt (2011)). Our measures of immigrant patenting activity agree with these survey findings. We build on this prior work by also documenting differences in knowledge diffusion and collaboration by immigrants and natives, since we link directly to the patent data itself. Other papers have focused on firm-level outcomes using changes in H1-B visa caps to estimate how marginal changes in immigration levels impact firm-level innovation (e.g., Doran et al. (2014)). Additional work has used state-level innovation measures (Hunt and Gauthier-Loiselle (2010); Chellaraj et al. (2005, 2008)). However, these approaches do not identify differences in productivity between individual immigrants and natives, and do not separate out spillover effects from direct output difference between natives and immigrants. Finally, some papers provide a historical perspective. Moser et al. (2014) shows that Jewish immigrants from Nazi Germany increased aggregate US innovation and raised the innovation output of native workers. Akcigit et al. (2017) links the now public-use 1880-1940 Censuses linked to patent records, showing that immigrants were disproportionate contributors to US innovation in the early 20th century. We add this literature by quantifying the contribution of high-skilled immigrants to overall US innovative output during the post-war era. Further, we are able to causally estimate a key mechanism through which high-skill immigrants create large, positive knowledge externalities on US-born inventors: human capital spillovers through patent collaborations.

Our paper also contributes to a literature studying immigrant assimilation and the effects of immigration on native employment outcomes. Several articles show evidence that immigrants are positively selected into developed countries (Abramitzky and Boustan (2017); Basilio et al. (2017); Abramitzky et al. (2014, 2012); Grogger and Hanson (2011, 2015). However, it is not clear whether this translates into higher productivity when in the United States due to potential assimilation issues. Most of these studies focus on wage outcomes, while we focus more directly on productivity as measured by patenting output. Indeed, since the US visa rules often give firms strong monopsony power over immigrant workers, wages may not be the best measure of productivity differences. Indeed, even in the early 1900s, Akcigit et al. (2017) find that immigrants produce more patents than native, but earn lower wages.

The remainder of the paper proceeds as follows. Section 2 describes the various data sources used in the analysis. Section 3 details our new empirical approach for identifying immigrant status and provides basic summary statistics. In Section 4, we characterize the immigrant share of US innovative output and explore life-cycle characteristics of immigrant and native productivity. Section 5 analyzes immigrant spillover effects. Section 7 concludes.

2 Data

We bring together data from multiple sources whose combination enables us to observe immigrant innovative productivity and explore how it compares to the innovative productivity of natives in the United States. Specifically, we combine patent data from the US Patent Office (USPTO) together with data provided by Infutor, which allows to identify immigrant status based on the combination of the first 5-digits of an individual's social security number (SSN) and their year of birth.

2.1 Infutor Database

The Infutor database provides the entire address history for more than 236 million US residents.[5] The address history generally dates back to 1990, although there are some individuals with entries dating back to the 1980s. For each individual, we have the exact street address at which the individual lived and the dates of residence. The data also provides the first and last name of the individual, as well as some demographic information such as year of birth and gender. Finally, in many cases the data provides the first 5-digits of the individual's social security number. This data was first described and made use of by Diamond et al. (forthcoming).

This data appears to be highly representative of the overall US adult population.[6] To examine the quality of the data, we use the address history provided and in each year map all individuals in the dataset to a US county. Using this mapping, we then create county-level population counts as measured by Infutor. We can compare these county-level populations with the population counts of over 18 years old individuals provided by the US census. Figure A.2 illustrates this relationship for the year 2000. We find that Infutor covers 78% of the overall adult US population as estimated by the US Census. Moreover, the data matches the cross-sectional distribution of US individuals across counties extremely well. The Infutor county-level population in 2000 explains 99% of the census county variation in population.

2.2 Patent Data

We obtain data on all U.S. patents granted from 1976 through 2015 directly from the United States Patent and Trademark Office (USPTO). The USPTO data provide information on the date a patent was applied for and ultimately granted, the individual(s) credited as the patent's inventor(s), the firm to which the patent was originally assigned, and other patents cited as prior work. From this, we can determine how many citations a granted patent receives in the future. The data also provides information on the technology class of the patent, as well as the city and state in which each inventor on the patent lives.[7]

[5]Infutor is a data aggregator of address data using many sources including phone books, magazine subscriptions, and credit header files.

[6]Infutor does not have any entries on one's address history as a child. In practice, people appear to enter the data at some point during their early to mid twenties.

[7]Note that these addresses are indeed the home addresses of the inventors, and not the addresses of the firms at which the inventors work.

One challenge the raw data presents is that it lacks consistent identifiers for patent inventors and firms over time. In order to identify inventors, we rely on a large-scale disambiguation effort provided by Balsmeier et al. (2015). Their algorithm combines inventor names, locations, co-authors, associated firms, and patent classifications to create an inventor identifier. Using this procedure thus gives us a panel of inventors, whereby in each year, we have data on any patents an inventor applied for (and was ultimately granted).

In the complete patent data-set, there are roughly 1.6 million unique inventors over the 1976-2015 time period residing in the U.S. It should be noted that we use the names of all individuals denoted as inventors in the patent documents, not just those who are assigned the intellectual property rights (i.e. the "self-assigned" holders of the patent rights). For example, if an inventor is working for a firm, it is usually the company who will be the awarded the patent rather than the employee herself. However, the employee will be still identified on the patent documents as the actual originating inventor, along with any co-authors. We therefore define a individual as a US-based inventor if he or she is named as such on the patent document and has a US address. We examine patenting between the years of 1976 to 2012 and we restrict our analysis to those inventors in the age range of 20 to 65 years old in any given year.

2.3 Measures of Inventor Productivity

To study differences in innovative output and productivity between immigrant and native inventors, we use a variety of patent-based measures that have been widely adopted over the past two decades (Jaffe and Trajtenberg (2002); Lanjouw et al. (1998)).[8] Our primary measure of the quantity of an individual's innovative output is the number of ultimately granted patents the individual applied for.

Our primary measure of the quality of a worker's innovative output is the number of citations the patents receive within some specified time frame. In general, we use a time window of three years since the grant date. Patent citations are important in patent filings since they serve as "property markers" delineating the scope of the granted claims. Furthermore, Hall et al. (2005) document that patent citations are a good measure of a patent's innovative quality and economic importance. Specifically, they find that an extra citation per patent boosts a firm's market value by 3%. Similarly, Kogan et al. (2017) find that patent's economic value is strongly correlated with its quality and scientific value as measured by patent citations.

One challenge in using patent citations as a standardized measure of innovative productivity is that citation rates vary considerably across technologies and across years. To address both of these issues, we normalize each patent's three year citation count by the average citation count for all other patents granted in the same year and 3-digit technology class. We call this measure Adjusted Citations. Finally, we construct a variable which we call Top Patents, which is a simple indicator variable equal to one if a patent was in the top 10% of patents from the same year and technology class in terms of citations received. This variable identifies a subset of highly influential patents

[8]More recent contributions include Lerner et al. (2011); Aghion et al. (2013); Seru (2014); Bernstein (2015).

granted within a technology class in a given year.

Finally, we additionally use a measure developed by Kogan et al. (2017) of the actual economic value generated by a patent. The measure is based on the stock market reaction to the announcement of the patent grant. Naturally, the manner in which this variable is constructed restricts the analysis to the sub-sample of patents assigned to publicly traded firms. Kogan et al. (2017) find that median economic value generated by a firm is substantial ($3.2 million in 1982 dollars).

2.4 Merging the Patent Data to Infutor

Our ultimate goal is to use the first five digits of the SSN and age information provided by Infutor to determine whether a US-based inventor is an immigrant or not. We therefore need to merge the patent data to the Infutor data. The feature of the patent data which allow us to do this is that if an inventor is ultimately granted a patent, we know the city and state in which the inventor was living when the patent was applied for, on top of her name. Since the Infutor database provides the entire address history of individuals dating back to the 1990s, we can then use name matching within a given city and year to merge the two datasets. This name matching follows an iterative process over multiple stages described in precise detail in Appendix **A**.[9] In the end, our procedure yields a total of roughly 915,000 matches, corresponding to a match rate of approximately 70% of all US-based inventors.

To assess the quality of the match, we explore selection on observables by comparing the innovative productivity of matched and unmatched inventors. In Table A.1 in the Appendix we find no strong selection effects associated with the matching. The average productivity of matched and unmatched inventors (as well as median and top 90%) is similar along the various productivity measures such as number of patents, total adjusted citations, number of top patents and total economic value created.

3 Identifying Immigrant Inventors

We use information regarding the first five digits of an individual's Social Security Number (SSN), in combination with information regarding the individual's age, to determine immigrant status. The essential idea is straightforward. The first five digits of the SSN pin down within a narrow range the year in which the number was assigned. When combined with information regarding the individual's birth year, we can determine how old the individual was upon being assigned the number. Since practically all US natives are assigned a SSN during their youth, those individuals who receive a SSN in their twenties or later are extremely likely to be immigrants. We apply this methodology to our merged data described in the previous section, thus allowing us to study the contribution of immigrants to US innovative output. Clearly this method will miss those who immigrated to the US prior to age 20. We investigate what share of immigrants we should expect to miss using using 2014 ACS data. We find that 17.1% of adults are foreign born, while 10.4% of adults are foreign

[9] Bernstein et al. (2018) follow a similar procedure in matching patent records to deeds records.

born and immigrated at age 20 or later, implying 39% of all immigrants in 2014 immigrated prior age 20. This number falls to 32% among college graduates and 19% among PhDs. This suggests we will classify some immigrants as natives, implying our analysis focuses on those who immigrate during adulthood.[10] A second issue is that we will miss illegal immigrants, as they would not have an SSN. However, this is likely less of an issue for high skilled immigrants who are inventors, since they would likely be employed in the formal sector.

Since our approach relies closely on the structure and precise assignment method of US Social Security numbers, we start by outlining the relevant history and institutional details of the SSN program. We then detail our exact approach of identifying immigrants using micro-level SSN and age information provided by Infutor. Finally, we perform several empirical tests to verify the validity of our immigrant classification methodology.

3.1 Institutional Details of SSN

The Social Security Number (SSN) was created in 1936 for the sole purpose of tracking the earnings of U.S. workers, so as to determine eligibility for Social Security benefits. By 1937, the Social Security Administration (SSA) estimated that it had issued 36.5 million SSNs, capturing the vast majority of the U.S. work force at that time. Since that time, use of the SSN has substantially expanded. In 1943, an executive order required federal agencies to use the SSN for the purpose of identifying individuals. In 1962, the IRS began using the SSN for federal tax reporting, effectively requiring an SSN to earn wages. In 1970, legislation required banks, credit unions, and securities dealers to obtain the SSNs of all customers, and in 1976 states were authorized to require an SSN for driver's licenses and vehicle registrations. Since its origination, the SSA has issued SSN numbers to more than 450 million individuals. Today, the SSN is used by both the government and the private sector as the chief means of identifying and gathering information about an individual. Practically all legal residents of the United States currently have a Social Security Number.

Since its establishment in 1936, and until 2011, Social Security numbers were assigned according to a specific formula.[11] The SSN could be divided into three parts:

$$\underbrace{XXX}_{\text{area number}} - \underbrace{XX}_{\text{group number}} - \underbrace{XXXX}_{\text{serial number}}$$

The first three digit numbers of the SSN, the area numbers, reflect a particular geographic region of the United States and were generally assigned based on the individual's place of residence. Groups of area numbers were allocated to each state based on the anticipated number of SSN issuances in that state.[12] Within each area number, the next two digits, the group numbers, were assigned

[10] Note that immigrants classified as natives are unlikely to affect the characteristics of the natives group given their particularly small fraction relative to the overall group.

[11] The Social Security Administration changed the structure of SSN numbers in 2011 to randomly assign all the parts of the SSN.

[12] If a state exhausted its possible area numbers, a new group of area numbers would be assigned to it. There are some special cases of area numbers. For example, area numbers from 700 to 728 were assigned to railroad workers

sequentially. A given area would assign the next group number in the line of succession after all of the possible serial numbers, i.e. the last four digits of the SSN, ranging from 0001 to 9999 had been exhausted.[13]

The sequential, formulaic nature of the assignment process implies that Social Security numbers with a particular combination of the first five digits were only assigned during a certain year(s). In fact, this information is available from the Social Security Administration (SSA) through the High Group List that they maintained up until 2011. Designed to enable the validation of issued SSNs and to prevent fraud, this data provides, for each area number, the month and year when a certain two digit group number began to be issued.[14]

3.2 Identifying Immigrants

Using this mapping between the first five digits of the SSN and assignment years, we can use our Infutor data to classify US-based individuals as either natives or immigrants. The key aspect of the Infutor data which allows for this is that, in many cases, the data has information on both an individual's SSN as well as her age.

Historically, SSNs were typically assigned at the age of 16 when individuals first entered the labor force, but as the SSN's usage and popularity grew due to the legislative initiatives described above, individuals began to receive an SSN at earlier and earlier ages.[15] Figure A.3 in the appendix shows the 25th, 50th and 75th percentiles of the age distribution of SSN assignees by assignment year, as measured by Infutor. Consistent with what we have described, all three percentiles of the age distribution are always under 20 years old and the median is always around 16 years old or below. Moreover, after 1960 the average age at which individuals receive their SSN begins to considerably decline.[16]

---

until 1963. Area numbers from 580 to 584, 586 and from 596 to 599 were assigned to American Samoa, Guam, the Philippines, Puerto Rico and U.S. Virgin Islands. Area numbers between 734 and 749 or between 773 and 899 were not assigned until 2011. Finally, no SSN can have an area number of 666 or 000. For more details, see Puckett (2009).

[13]Group numbers were assigned in a non-consecutive order: first odd-numbers from 01 to 09, second even numbers from 10 to 98, third even numbers from 02 to 08, and finally odd numbers from 11 to 99. We encoded the group number to a sequential order from 01 to 99, so, for example, encoded group number 02 and 03 corresponds to SSN group 03 and 05 respectively. That is, our encoded group numbers reflect the true position in the line of succession, rather then the actual SSN group number. This simplifies the graphical illustrations discussed in the text.

[14]The High Group list is available on the ssa.gov official website. Its publication ended in 2011 due to the implementation of SSN Randomization. Since the historical information on Group Number assignment years, however, is available on the SSA website from 2003 only, we use an alternative data provider, *www.ssn-verify.com*, also based on the historical High Group Lists, to collect group number assignment years dating back to 1950. We verify the accuracy of the reported assignment year by checking that within each group number, the assignment year corresponds to the highest year of birth within the cohort that has that SSN (that is, reflecting individuals that were just born). This data provides us with information on assignment years between 1951 and 2011. Before 1950 we imputed the assignment year by simply adding 16 years to the most frequent year of birth within group number. This assumes that most people got their SSNs when they were 16 years old before 1950. We show that this imputation is valid because there is no discontinuity of encoded group numbers sequence around 1950 for each area number (A.4).

[15]By 2006, more than 90% of SSNs were being assigned at birth.

[16]In 1986, as part of the Tax Reform Act, the IRS began to require an SSN for all dependents older than age 5 reported on a tax return. The law further required that student loan applicants submit their SSN as a condition of eligibility. In 1987 the "Enumeration at Birth" (EaB) program started, which allowed parents of newborns to apply for an SSN as part of the birth registration process.

Given these considerations, we classify as an immigrant all individuals in the Infutor data who are more than twenty years old when assigned an SSN.[17] We also explore alternative, more conservative classifications of immigrants, requiring gaps of 21 to 25 years between the SSN assignment year and the individual's birth year. Our results are robust to these alternative classifications. In the next subsection, we explore how representative our classification of immigrants is when compared to three different sources of aggregate statistics of immigrants in the United States.

3.3 Validation Tests

We begin by comparing the proportion of county-level immigrants based on the entire Infutor data-set and our new classification methodology to the proportion of foreign born individuals at the county level in the 2000 Census.[18] To do so, we first geo-code individuals in the Infutor data-set to US counties based on their exact 2000 street address. From this mapping and our immigrant classification procedure, we then calculate the immigrant proportion of the 2000 county population. We perform this calculation several times as we apply different SSN assignment cutoffs between ages of 20 to 25. We finally run regressions of the proportion of foreign born individuals as measured by the Census on our constructed measures. In each regression, we use the 2000 population size as reported by the 2000 Census as weights.

Figure A.5 in the Appendix reports the R^2 of these regressions. The x-axis denotes the minimum gap between the SSN assignment year and birth year that is required to classify an individual as an immigrant. Comfortingly, all of our specifications produce R^2 of approximately 90%. This test illustrates that our immigrant classification procedure captures well the cross-sectional variation in immigrant shares across US counties. Figure A.6 provides binscatters of these regressions. While we match the cross-sectional variation extremely well, these results also illustrate that, on average, the proportion of foreign born in a county according to the 2000 census is slightly above 1.5 times the proportion of immigrants predicted by our method. This is expected, however, because the Infutor data only contains adults and legal immigrants, while the CENSUS counts all age groups as well as undocumented immigrants.[19]

To explore whether our immigrant classification method can do even a better job in explaining variations of immigrants shares when we focus on adults only we use the ACS, which allows us to not only incorporate individuals age but also, importantly, identify the age in which immigrants

[17] We classify all individuals that have a SSN that is either an ITIN or belongs to Enumeration at Entry program as immigrants as well. Summarizing, if we sum all the special cases that we don't account for in the immigrant classification (U.S. territories, not issued areas, not valid areas, group number 00, railroad and not issued groups) they represent 0.83% of the Infutor data.

[18] The 2010 CENSUS does not have the proportion of immigrants at the county level.

[19] In Figure A.7 in the Appendix we plot the combined R^2 and regression coefficients for age thresholds between 10 years old to 30. As expected, the lower the age threshold, the lower the regression coefficient, implying that the share of foreigners, based on this classification is increasing, as we classify younger and younger individuals as immigrants. However, it is important to note the changes in the R^2. As we approach the age threshold of 20, our ability to explain variations in immigrants across counties increases, and stabilizes around the age of 20, consistent with the notion that around that age threshold we are indeed able to separate immigrants and natives based on the age in which they received their social security number.

arrived to the US. In principle, this allows us to identify in the ACS exactly those immigrants we propose to identify in Infutor. To have a representative sample at each age, we use the ACS at the state level rather than at the county level and calculate the proportion of the population that is both foreign born and immigrated after they had reached 20 years of age. Similar to what we did previously, we then regress the proportion of the state population of a certain age that is both foreign born and immigrated after the age of 20, as reported by the ACS, against the same statistic constructed through Infutor.

Figure A.8 illustrates the fit of these regressions through binscatters using the 2005 ACS for several adult age groups. For example, panel (a) provides the binscatter for adults in ages of 40-44. The R^2 in that case if 94%, and consistent with the notion that we have a more comparable group now, explains better the cross-section variation of immigrant proportion. Moreover, it is also useful to note that the under-representation of immigrants declines, again, consistent with the fact that we no longer pool immigrants that arrived as kids to the US. We find similar results when we focus on age groups 45-49, 50-54, and 55-59, when the R^2 ranges between 94%-97%. The ACS shows approximately 30% more immigrants than our data, this is expected because our immigrant classification does not account for illegal immigrants. Indeed, the Department of Homeland Security estimates that 34% of immigrants were illegal in 2014. This matches very closely with the 30% under count of immigrants in Infutor, further validating our methods.

3.4 Summary Statistics

Table 1 provides summary statistics at both the inventor level and the patent level for our final sample. We first see that the productivity distribution for inventors is highly right-skewed. The median inventor has two patents, four citations, and approximately 1 adjusted citation over the course of a career. The median inventor also generates only \$20.5 million of economic value, as measured by KPSS stock price reaction measure, and no top patents. The mean inventor, in contrast, has 4.41 total patents, 21.88 total citations, 5.82 adjusted citations, and 0.88 top patents. Most significantly, the mean inventor is associated with patents generating \$98.9 million of economic value.

This right-skewness is also apparent at the patent level. The median patent has 2 citations, 0.52 adjusted citations, and generates \$7.2 million in economic value. The mean patent has 4.47 citations, 1.22 adjusted citations, and generates \$18.42 million of economic value. The table also reports that the mean age of an inventor filing a patent is 45 years (median is 44).

Finally, Table 1 provides some basic summary information on the demographics of inventors in our sample. Ten percent of the inventors in our sample are female and 16 percent of the inventors are immigrants to the United States.

4 Results

In this section, we explore the innovative contributions and patterns of US immigrant inventors over recent decades. We begin by exploring the contribution of immigrants to total US innovative output, relative to their share of total US-based inventors. We then examine the innovative productivity of immigrants over their life-cycle, and compare these patterns to US natives. Next, we explore the role of immigrant inventors in fostering the global diffusion of knowledge and, finally, we analyze the extent to which immigrants appear to assimilate into the broader US inventor pool over time.

4.1 Immigrants' Share of Innovation

Figure 1 illustrates that 16% of US-based inventors immigrated to the United States when they were at least 20 years old. This number is line with statistics provided by the 2016 ACS. According to the ACS, 16% of workers in STEM occupations were immigrants who immigrated at age 20 or later.[20]

Given that we find 16% of inventors in our sample are immigrants, the next natural question is to determine the overall share of US innovative output between the years of 1976 to 2012 was produced by immigrants. To calculate the relative share of immigrants in innovative production, however, we need to account for the fact that some patents are produced in teams. Therefore, to calculate an individual inventor's output, we divide each patenting variable of interest by the size of the team associated with that patent. For example, if four inventors are listed on a patent, we assign each inventor a quarter of a patent, and divide the number of citations and patent market value by four.

We find that immigrants account for approximately 22% of all patents produced over the time period of our sample. Remarkably, this represents more than 40% increase relative to their share of the US-based inventor population. One possibility, though, is that immigrants might be producing more patents of lower quality than their US native counterparts. We find that this is not the case. The fraction of raw future citations attributed to immigrants in our sample is again roughly 22%, suggesting that the higher production of patents by immigrants is not coming at the cost of the lower quality. Still, yet another concern is that immigrants may select into technologies that have higher citation rates, which could account for these results. Looking at adjusted citations, however, in which we scale citation rates by the average citations of all patents granted in the same year and technology class, we find that the contribution of immigrants is if anything slightly higher, accounting for 24% of the total. Similarly, when we focus on the production of top patents, those patents that are at the top 10% of citations within a technology class and year, we find a similar pattern, with immigrants generating roughly 24% of top patents in our sample period.

Finally, we explore the share of total economic value that immigrants have generated over the last four decades. To do so, we rely on the Kogan et al. (2017) measure that captures stock market reaction to patent grants. We find that the immigrants have generated 25% of the aggregate

[20] STEM occupation defined as engineers, mathematical and computer scientists, natural scientists, and physicians.

economic value created by patents in publicly traded companies between the years of 1976-2012. One might worry that this last result is driven by selection, to the extent that immigrants are more likely to work in publicly traded firms than US natives. Again, this does not appear to be the case. We find that immigrants account for 17% of the inventor workforce. Hence, immigrants are not dis-proportionally sorting into publicly traded firms. Relative to their share of the total number of inventors working in publicly traded firms, the economic value created by immigrants reflects an increase of 47%.

We finally explore whether the contribution of immigrants to innovation is concentrated in particular technology categories. In Figure 2, we construct the relative contribution of immigrants across six technology categories. Immigrants account for about 25% of patents among four main technological categories that were emerging during our sample period: Computers and Communications, Drugs and Medical, Electronics, and Chemical technologies. In contrast, the presence of immigrants seem to be lower at about 15% in more traditional technologies such as the "Mechanical" category that involves Metal working, Transportation, Engines, and the "Other" category that includes various technologies related to Heating, Agriculture, Furniture, among others.

4.2 Inventor Productivity over the Life-Cycle

The previous section illustrates the disproportionate contribution of immigrants to overall US innovative output, relative to their share of in the US-based inventor population. In this section, we begin to unpack the source of these differences, exploring the innovative productivity of both immigrants and US natives over the life-cycle. To do so, we compile for each individual her patenting activity throughout the span of her career.

Panel (a) of Figure 3 illustrates the life-cycle innovative productivity of native and immigrant inventors as measured by the annualized number of patents. For both populations, we see that, on average, the number of patents per year increases rapidly during the 30s, peaking in the late 30s, and then declines slowly into one's 40s and 50s. While the innovative productivity of natives and immigrants follow similar trajectories early in the life-cycle, the two populations diverge when reaching the peak of innovative productivity, with immigrants significantly more productive than natives. At its peak, the gap amounts to 50% higher productivity of immigrants. The gap, while somewhat declining, continues to persist throughout the rest of their careers.

While the number of patents may not necessarily capture the quality of the underlying innovation, a similar pattern is apparent in Panel (b) of Figure 3, in which we measure innovative productivity according to the annualized sum of citation-adjusted number of patents. For both immigrants and natives, we find an inverse U-shape pattern of inventor productivity, but immigrants become significantly more productive than natives in terms of adjusted citations from mid-30s and onward. At its peak, based on this measure, the gap suggests that immigrants are almost twice as productive as natives. These patterns are also confirmed in Panels (c) and (d) of Figure 3, which respectively provide measures of the annualized production of top patents and total economic value generated.

The inverse U-shape productivity of native and immigrant inventors is consistent with a large literature exploring the relationship between age and scientific contributions. See Jones et al. (2014b) for a survey. This research consistently finds that performance peaks in middle age: the career life-cycle begins with a training period in which major creative output is absent, followed by a rapid rise in output to a peak, often in the late 30s or early 40s, and finally ending with a subsequent slow decline in output through one's later years (e.g., Lehman (1953); Zuckerman (1977); Simonton (1991b,a); Jones (2010), among others). These patterns are consistent with theoretical models of human capital accumulation in which researchers invest in human capital at early ages, and, in so doing, spend less time in active scientific production. Consequently, skill is increasing sharply over time but is, initially, not directed towards output. Eventually, researchers transition to active innovative careers (Becker (1964); Ben-Porath (1967); McDowell (1982); Levin and Stephan (1991); Stephan and Levin (1993); Oster and Hamermesh (1998)). Researchers also surely benefit from learning-by-doing (Arrow (1962)), which provides yet another source of increasing output overtime. Such models may explain the low productivity of immigrants and natives early on in the life-cycle, but do not account for the differences in productivity between immigrants and natives around the peak productivity point.

4.3 Cohort Effects and Differential Sorting

In this section, we consider a variety of potential explanations for the life-cycle differences in productivity between immigrants and natives, including cohort effects and differential sorting across industries and space. First, Jones (2009, 2010); Jones and Weinberg (2011) emphasizes that the age-output profile within fields is not fixed but has actually changed quite dramatically over time. In line with a "burden of knowledge" view of the innovative process, he observes that the quantity of precursor scientific and technological knowledge has expanded substantially over time, leading high quality, significant technological contributions to shift towards later ages. This implies that the life-cycle pattern of productivity might depend on birth cohort. A potential concern which arises from this, then, is that our results on the gap between immigrant and native productivity could be driven by differences between immigrants and natives in the distribution of birth years.

Another concern is that immigrants may simply work in different technology classes than natives. Then, to the extent that it is easier to innovate in certain technology classes, certain technology classes have more impactful innovations, or the burden of knowledge is lower in some technology classes, we would find differences in the innovative output of immigrants versus natives over their life-cycles. A related concern is immigrant inventors may be deferentially sorted into different regions in the United States. To the extent that immigrants, often thought to be more mobile than natives, are more likely to settle in innovation hubs, i.e. regions which foster innovative productivity through local agglomeration spillovers, such geographic sorting might account for the measured productivity gaps. See, for example, Marshall (1890); Jaffe (1989); Audretsch and Feldman (1996); Ellison et al. (2007), among others. Indeed, in 2005, 13.2% of immigrant inventors lived in Santa Clara County, i.e. Silicon Valley, while only 4.4% of native inventors did so.

We explore the importance of these channels in a regression setting in Table 2. In panel (a) we explore these effects on the annual number of patents. We start in column (1) by simply controlling for year of application fixed effects. Immigrants seem to produce on average 0.106 higher number of patents per year, and the effect is highly statistically significant. In column (2) we add year of birth fixed effects, which account for variations across cohorts in the time required for training and human capital accumulation to reach the knowledge frontier, as discussed by Jones (2009, 2010); Jones and Weinberg (2011). We find that the coefficient remains unchanged. In column (3), we also add county fixed effects, comparing individuals who reside in the same region, and thus likely benefiting from the same local knowledge spillovers and agglomeration externalities. The innovation gap between immigrants and inventors does decline, but is still positive and highly statistically significant at 0.089 patents per year. In column (4), we also allow for sorting across technology classes by including county by technology class fixed effects in addition to year fixed effects and YOB fixed effects. The results are largely unchanged. In column (5), we allow for the possibility that local county agglomeration benefits vary over time and include county by year fixed effects. Finally, in our most stringent specification, we include county by technology class by year fixed effects in addition to YOB fixed effects. There is still a substantial productivity gap between immigrants and natives. Immigrants produce .0786 more patents per year, even when accounting for these sources of differential sorting.

These results suggest that differential sorting, particular regional sorting, can explain some of the productivity gap between immigrants and natives, but still cannot account for the large majority of the difference. In general, regional sorting appears to account for 35% of the productivity gap.

In panel (b) we explore the effect of these channels on annual citation-adjusted number of patents, in panel (c) we explore the effect on annual economic value, and finally in panel (d) we explore annual production of top patents. In all of these measures we find that while the gap seem to decrease, between immigrants and natives, once we hold these differential sorting factors fixed, it nevertheless remains quite large and highly statistically significant. Specifically, immigrants produce 0.109 more citation adjusted number of patents, $0.913 million more in economic value, and 0.022 more top patents.

In the Appendix we also explore whether the inverse U-shape of the innovation production function of immigrants and natives still remain when we add such controls, as well as whether at the peak of the career immigrants still remain significantly more productive. We repeat figure 3, but focus either on inventors that issued their first patent in the 1990s, or only on inventors that were born in the 1970s. The first empirical exercise fixes the transition of individuals from human capital accumulation to the pursuit of research, while the second empirical exercise ensures that we compare individuals that had similar time to accumulate human capital over the life cycle. The results are illustrated in Figures A.9 and A.10 in the Appendix, and portray similar findings.[21] Moreover, we

[21] Another advantage of these empirical tests arise from the particular procedure we use to construct our sample. As we noted previously, since the Infutor database coverage becomes significantly better during the beginning of the 1990s, we are less likely to capture individuals that stopped patenting before the 1990s, thus introducing selection bias when measuring patenting output in the 1980s. Comparing individuals that first started to patent in the 1990s,

also explore these patterns when holding individual fixed effects. Specifically, in Figure A.11 in the Appendix, we report the coefficients that interact age with an immigrant status dummy, while holding year and individual fixed effects. This specification absorbs both cohort fixed effects, skill and specialization of inventors that are time invariant. The results nevertheless remain unchanged.

4.4 Immigrant Integration into Global Knowledge Market

Do immigrant inventors bring unique knowledge to US innovation markets? Some theories of human capital accumulation and longstanding conceptions of creativity define a cognitive process where new ideas are seen as novel combinations of existing material (Usher (1954); Becker (1982); Weitzman (1998)). One potential benefit of immigration to the United States, therefore, is the importation of global knowledge and the integration of foreign ideas with US-based ideas. Indeed, immigrants may be trained and exposed to vastly different types of technologies and ideas in their origin countries, relative to the United States. This suggests that immigrants may be uniquely positioned to explore novel combinations of knowledge acquired in their home countries, together with technologies to which they are exposed in the U.S. In fact, in surveys of Silicon Valley, 82% of Chinese and Indian immigrant scientists and engineers report exchanging technical information with their respective nations (Saxenian (2002); Saxenian et al. (2002)).

To explore the extent in which immigrants are more likely to import and integrate foreign technologies, we further explore the details of US-based innovative output, particularly the reliance on foreign technologies and collaboration with foreign inventors. Our results are reported in Figure 4. In Panel (a) we explore the extent to which immigrants and natives rely on non-US technologies. To do so, we calculate for each patent, the share of backward citations of patents that were issued outside the United States. We present the share of foreign backward citations separately for natives and immigrants over their life-cycle. As Panel (a) illustrates, immigrants are significantly more likely to rely on foreign technologies in their patent production, when the gap amounts to more than 10%. In Panel (b), we find that immigrants are significantly more likely to collaborate with foreign inventors, relative to native inventors. Specifically, on average, immigrants collaborate with at least one foreign inventor in 16% of their patents, in contrast to 9% of native inventors.

Finally, in Panel (c), we provide an additional measure that explores the extent to which immigrants are integrated in global innovation markets by exploring how likely foreign inventors are to cite immigrant patents relative to native patents. As expected, we find that immigrants' patents are more likely to be cited by foreign inventors, illustrating the fact that immigrant innovation not only disproportionately draws from foreign markets, but is also disproportionately visible to foreign markets. All of this evidence together supports the view that immigration to the United States fosters the global diffusion of knowledge and the integration of foreign and US ideas.

The findings that immigrants to the US remain integrated with global knowledge markets, and

as reported in Figure A.9 in the Appendix, ensures that our results are not driven by sample selection associated with our sample construction procedure. Similarly, comparing individuals that were born in the 1970s, as reported in Figure A.10 in the Appendix, once again suggests that the first patent is very likely to take place during the 1990s, providing yet another useful test of the robustness of our results.

contribute to the cross-border diffusion of technologies is consistent with the idea of a worldwide technology frontier, where new ideas and innovations travel quickly to all countries. However, knowledge transfer may be more complicated than simply sharing blueprints, process designs, or journal articles due to the often tacit knowledge associated with new innovations shapes. In that regard, immigrants contribute uniquely to the transfer of such technologies.

Finally, it is interesting to note that the gap between immigrant and native inventors in terms of the tendency to collaborate with foreign inventors, or to be cited by foreign inventors is declining over time. The result may be driven by increasing assimilation of immigrant inventors over time. We directly explore this question in the following subsection.

4.5 Assimilation of Immigrants in the US

We might expect that differences in language and culture may limit the ability of immigrants to collaborate and integrate into the local labor market (see Borjas (2014) for a formalization of this idea). Alternatively, immigrants' investments in US-specific skills may have limited effect on collaboration with native inventors if immigrants face discrimination in local labor markets (Moser, 2012).[22] Assimilation difficulties may suggest that immigrants may be more inclined to either work in seclusion, or alternatively may be less inclined to work with native inventors. The extent to which immigrants collaborate with native inventors may have important implications for the spillovers and the indirect contribution of immigrants to US innovation.

The patent data provides a unique glimpse into the assimilation of immigrants into the US labor market over time, as patent application documents provide information on an inventor's collaborators.

In Panel (a) of Figure 5, we explore whether immigrants are more likely to work in seclusion, or less likely to collaborate, with US inventors over time. We do so by constructing the number of unique co-authors that appear on an inventor's patent applications in a given year, as a proxy for the number of inventors that an individual collaborates with. As Panel (a) shows, in their early years, natives and immigrants exhibit similar patterns, in terms of the number of unique inventors with which they collaborate. However, immigrants seem to work with a higher number of individuals during their 40s and 50s, consistent with their higher productivity in those years. We find similar results in panel (b) when focusing only on co-authors that are based in the US.

We next explore the extent to which immigrants work with other immigrants and the extent to which they collaborate with US natives. If assimilation requires cultural adaptation, and acquisition of US-specific skills, we anticipate that over time we may see a gradual increase in the tendency of immigrants to collaborate with natives. Indeed, we find patterns that are very consistent with this hypothesis. In Panel (c) of Figure 5, we calculate the share of unique co-authors that are foreign born. Among natives, we see that the share of immigrant collaborators is fairly fixed and

[22] Moser (2012) exploits a change in attitudes toward a particular immigrant group—German Americans after the outbreak of World War I—to evaluate the effect of discrimination on immigrants' economic opportunities. She shows that, during (but not before) the war, men of German ancestry were more likely to be excluded from seats on the New York Stock Exchange.

equal to roughly 7% over their life-cycle. In contrast, for immigrants, early on in their careers, the share of unique immigrant co-authors is roughly 17% (more than twice the share of natives). However, unlike for natives, we also see a gradual decline over time in the propensity of immigrants to work with other foreign-born inventors. Again, as illustrated in panel (d), we find similar patterns when focusing only on collaborators who are based in the US. This gradual decline in the share of immigrant collaborators may suggest that immigrants increasingly assimilate over time, although, the gap never closes and even towards the end of the career, immigrants are still more likely to collaborate with other immigrants.

5 Team-Specific Capital

Thus far, we have established that the innovative productivity of immigrants differs significantly when compared to that of natives. In particular, immigrants seem to be more productive over the life-cycle of their career and to be more integrated into global knowledge markets, facilitating the diffusion of ideas. Moreover, immigrants also appear to work more with other immigrants, although this effect seem to decline over time, suggesting that immigrants gradually assimilate into the local labor market.

In this section, we explore yet another potential difference between immigrant and native inventors, in the form of their contribution to team-specific capital (Jaravel et al., 2018). Specifically, we address the extent to which natives and immigrants impact the productivity of their collaborators. Such positive effects may reflect, for example, skill complementarities, as well synergies of experience and knowledge which might be difficult to construct or achieve otherwise.

Measuring any given individual's contribution to team specific capital is challenged by the endogenous creation and ending of collaborative research efforts. The ideal research design, therefore, is to find situations in which the collaboration between two patent inventors exogenously ends, and then study if there is any significant and long lasting impact on the careers of the collaborators. For our purposes, we are particularly interested in whether such disruptions differ across immigrants and natives, that is, whether immigrants or natives yield a greater productivity boost to their co-authors.

To construct causal estimates, our identification strategy exploits the pre-mature deaths of inventors, defined as deaths that occur before or at the age of 60, as a source of exogenous variation in collaborative networks. This form of identification strategy is becoming increasingly common in the literature.[23] We primarily follow Jaravel et al. (2018), in which the causal effect is identified through a difference-in-differences research design using a control group of patent inventors whose co-inventors did not pass away, but who are otherwise similar to the inventors who experienced the premature death of a co-inventor. We then compare the relative impact of a pre-mature death of an immigrant on co-authors with that of a native to estimate their respective spillover effects.

[23] See, for example, Jones and Olken (2005); Bennedsen et al. (2008); Azoulay et al. (2010); Nguyen and Nielsen (2010); Oettl (2012); Becker and Hvide (2013); Fadlon and Nielsen (2015); Isen (2013); Jaravel et al. (2018).

In the next subsections, we describe the data construction and the compilation of the matched co-author sample. We then describe the empirical specifications we use to identify the causal contributions of immigrant and native inventors to team-specific innovative capital.

5.1 Data Construction

We first identify 15,032 deceased inventors that were granted a patent before their death. Information on the year of death and age at death is available from the Social Security Death Master File (DMF), which is a database file made available by the United States Social Security Administration (SSA).[24] It contains information on all Social Security numbers that have been retired since 1962 due to death of the individual. In 2009, the file contained information on over 83 million deaths. We only include inventors that are present in our Infutor sample so that their immigrant status can be determined.

As in Jaravel et al. (2018), we construct a group of "placebo deceased" inventors who appear similar to the deceased inventors on various dimensions, who did not pass away, and who are not coauthors of the deceased inventors. Specifically, we match placebo deceased inventors based on immigrant status, the age at (real or placebo) death, the cumulative number of patent applications at the time of (real or placebo) death, the calendar year of (real or placebo) death, and finally the cumulative number of coauthors at the time of (real or placebo) death, grouped into ventiles. We find matches to all 15,023 deceased inventors using this procedure.

Next, we restrict the sample in the following ways. First, when there are multiple matches to deceased inventors, we randomly select one match to get a sample of one-to-one exact matches. Next, we restrict our sample to only those inventors who died at the age of 60 or earlier. The goal of this restriction is to primarily capture only unexpected, sudden deaths. Older individuals may have prolonged periods of ill health prior to death, leading to pre-trends in the analysis. By plotting the dynamics of the effects below, we will show that there indeed does not appear to be any pre-death deterioration in the productivity of the deceased inventor co-authors. Applying these restrictions, there are 9,405 real deceased inventors and the same number of placebo deceased inventors.

In Panel (a) of Table 3 we provide summary statistics for the real deceased and matched placebo deceased inventors. By construction, real deceased and placebo deceased inventors are perfectly balanced on age, year of death, immigrant status, and cumulative patents. At the time of death, the deceased is, on average, 51.5 years old and has filed an average of 2.8 patents. Nine percent of the deceased sample are immigrants. Since we match also on accumulated number of co-author ventiles pre-death, real and placebo deceased are very balanced on that dimension as well, with 5.7 and 5.5 co-authors respectively..

Panel (a) of Table 3 also shows that real deceased and placebo deceased are well-balanced on other measures of patenting productivity, despite not explicitly matching on these variables, providing further validation of our procedure. For example, real deceased inventors have an average of 3.3 total adjusted citations, have 0.42 top patents, have generated an average of $27.8 million of

[24]We accessed a public-use copy of the Social Security Death Master File courtesy of SSDMF.INFO.

economic values, worked on average with a team size of 1.9 collaborators. These statistics for the placebo deceased are, respectively, 3.8 adjusted citations, 0.48 top patents, $27.2 of economic value, and a teamsize of 1.9 collaborators.

Finally, we build the entire co-author network of collaborators prior to the death for each of the real and placebo deceased inventors. This yields 27,393 co-inventors of the placebo deceased, whom we we refer to as placebo survivor coauthors, and 27,796 co-inventors of the real deceased inventors, whom we refer to as real survivor inventors. In some analyses we rely on a different sample as well in which we further restrict all of these co-authors to those inventors that are matched to the Infutor sample, so that we can identify the immigration status of the deceased inventor co-authors (placebo and real). Applying this restriction leaves us with 17,477 placebo survivor co-authors and 16,829 real survivor co-authors. Our main results, however, are robust to applying this restriction, and yield similar results in both samples.

Panel (b) of Table 3 provides summary information on the real and placebo co-authors. We once again find that, despite not explicitly matching on the characteristics of co-authors or the strength of collaboration, the sample of real and placebo surviving co-authors is quite balanced. The surviving co-authors of real deceased are, on average, 51.3 years old. Sixteen percent are immigrants and 10 percent are female. Placebo co-authors are, on average, 49.3 years old, with 17 percent immigrants and 9 percent female. Real surviving co-authors co-patented, on average, 2.0 patents with the deceased prior to death. They have, on average, filed 10.6 cumulative patents, 2.1 top patents, and received 15.0 total adjusted citations. Placebo surviving co-authors are very similar. On average, they have co-patented 2.1 innovations with the deceased, filed 10.0 cumulative patents, 2.1 top patents, and received 15.8 total adjusted citations.[25] In Panel (c) we also compare the distribution of patents across technologies for real and placebo deceased inventors as well as their collaborators. Overall, the distributions seem to be quite balanced across both populations.

5.2 Research Design

Our goal is to estimate the causal effect of an inventor's death on the innovative productivity of real survivor coauthors, and compare the magnitude of this effect between immigrant and native inventors. Naturally, the productivity of co-authors of deceased inventors may have a different innovative trajectory than the full population of inventors. For this reason, we use as a control group the co-inventors of placebo deceased inventors described in the previous sub-section. Moreover, we need to ensure that inventor deaths are exogenous to collaboration patterns. Indeed, as we will show below, we find no statistically significant pre-trends, with the estimated causal effects of co-inventor death becoming statistically significant only after the year of death.

[25] One perhaps surprising aspect of Panel (b), Table 3 is how productive the real and placebo surviving co-authors are relative to the average inventor in the full sample. In fact, this is very consistent with Jaravel et al. (2018). As that paper notes, this is due to selection. More productive inventors, i.e. those who have generated a lot of patents, are more likely to experience the (real or placebo) death of a collaborator. Indeed, this selection is exactly why it would not be appropriate to use the full sample of inventors as a control group and why, instead, we use the placebo co-author survivors.

Our identification strategy is similar to that of Jaravel et al. (2018). To study the dynamics of the effect and test for pre-event trends, we use a full set of leads and lags around co-inventor death specifically for real survivor inventors (L_{it}^{real}) as well as a full set of leads and lags that both real and placebo survivor inventors (L_{it}^{all}) within each matched pair m of real and placebo dying inventor.[26] This allows for arbitrary trends within the set of surviving inventors among each matched pair of real and placebo dying inventors. This additional controls give us more power. Specifically, we estimate the following OLS specification:

$$Y_{it} = \sum_{k=-9}^{9} \beta_k^{real} \mathbb{1}_{L_{it}^{real}=k} + \sum_{k=-9}^{9} \beta_{mk}^{all} \mathbb{1}_{L_{it}^{all}=k} + \alpha_i + \epsilon_{it} \quad (1)$$

The effects of interest are denoted β_k^{real}, where k denotes time relative to death. These estimates reflect the causal effect of co-inventor death on the outcome of interest k years around death. Note that the joint dynamics around death for both real and placebo survivors is captured by β_{km}^{all}. We also include individual fixed effects (α_i), absorbing individual time-invariant characteristics.

To summarize the results and discuss magnitudes, we employ a second specification that relies on an indicator variable that turns to one after the real death of the inventor ($AfterDeath_{it}^{real}$), but maintaining the same controls as equation 1. Thus, β^{real} gives the average causal effect of death on collaborators. We also estimate this second specification by OLS:

$$Y_{it} = \beta^{real} AfterDeath_{it}^{real} + \sum_{k=-9}^{9} \beta_{mk}^{all} \mathbb{1}_{L_{it}^{all}=k} + \alpha_i + \epsilon_{it} \quad (2)$$

Note that this model once again includes year and individual fixed effects. We estimate equations (1) and (2) for the full sample of real and placebo survivors, and then separately for real and placebo survivors of immigrant and native inventors. Finally, we estimate separately the effect of immigrants pre-mature deaths on immigrant co-authors and native co-authors, and repeat the same empirical exercise for natives' pre-mature deaths. In all analysis, we cluster standard errors at the deceased inventor level.

5.3 Results

We examine four outcomes: number of patents, patents in the top 10% of citations in their technology class (Top Patents), weighted number of patents by adjusted citations, and economic value. Our results from equation (2) are reported in Table 4, which reports β^{real}. For all inventors, we see economically meaningful and statistically significant declines in innovative productivity across all measures, except adjusted citations. Moreover, across all measures, Moreover, across all four measures of innovative productivity, we find that that co-inventors of immigrants face a larger decline in the years subsequent to a collaborators death, suggesting that the causal effect of an immigrant inventor death on his or her team is larger than that of a native inventor.

[26] We only include data within event years -9 to 9 in the regression.

We first focus on the annual number of patents produced. In column (1) of panel (a), we provide the estimate for all inventors, regardless of whether the deceased inventors are immigrants or natives. The dynamic treatment effects around year of death are plotted in Figure 6. The coefficient β^{real} equals to -0.0595 and is highly statistically significant. Thus, relative to placebo co-authors, those inventors who experience the real death of a collaborator are significantly less productive. To interpret these magnitudes of the treatment effect, we quantify the percent change in the outcome, relative to the expected mean outcome of the treatment group, had they not been treated.[27] Relative to this expected mean (reported in Table 4), the treatment effect implies that a deceased inventor lead collaborators to produce 12% lower patenting output. In column (2), we explore the effect of a premature death of an immigrant. We find that the decline in the number of patents of co-inventors is significantly larger. The coefficient equals -0.195 and is again highly statistically significant, implying a 26% decline in patenting. In contrast, in column (3) we focus on the causal effect of pre-mature death of natives, and find that the magnitude of the decline in productivity of co-inventor, as measured by number of patents, while still statistically significant, is only 9%.

In columns (4) to (6) of Table 4, we focus on the *Top patents* measure and find similar results. The dynamic treatment effects around year of death are plotted in Figure 7. As shown in column (4), for all inventors, we find a statistically significant coeffieicient of -0.0220, which is equivalent to a 15% decline in the production of top patents following a collaborator death. Again, the effect is significantly higher for immigrants. Specifically, as reported in columns (5) and (6), immigrant inventor death leads to a decline of almost 27% in the number of top patents filed by co-authors, while the effect is only 11% for natives. In Panel B we explore two additional dimensions of innovative productivity, the number of patents weighted by adjusted number of citations and the economic value of patents. We find similar patterns. The death of a collaborator leads to a decline in innovative productivity, but the effect of a death of an immigrant is significantly larger. The dynamic treatment effects around year of death for these outcomes are plotted in Figures 6 and 7. In our adjust citation measure, it does appear that the treatment begings to take effect a few years before the actual death. This is likely due to our our adjusted citations metric is calculated. We measure the number of citations in the first three years after the patent is *granted*. Since the grant date is usually two to three years after the application date, one could imagine the death of the inventor impacts citations post death.

Finally, Table 5 reports the results of equation (2) but exploring separately the effects of immigrant and native deaths on immigrant and native co-authors. In this analysis, we must be able

[27] Specifically, we calculate the expected mean counterfactual for the treatment group by estimating a simple regression specification that is standard in the diff-in-diff framework: $Y_{it} = \beta_0 + \beta_1 Treat + \beta_2 Post_t + \beta_3 Treat_t Post_t + \epsilon_{it}$. The estimated outcome of the treatment group, absent treatment is: $\beta_0 + \beta_1 + \beta_2$. This simplified regression removes the individuals fixed effects and replaces them with a dummy for being in the treatment and replaces the calendar year fixed effects with the dummy for being in the post period (after either placebo or real death). The allows us to quantify the average outcome in the post period for the treatment, absent treated by essentially averaging the individuals fixed effects together into the treatment dummy, and average the calendar year FEs into the post dummy.

to identify the immigrant status no only of the dying inventor, but also of his collaborators. This restricts our sample of analysis because not all of our collaborators match to our Infutor data and not all of the matches have an SSN listed in infutor. To mitigate the effects of this restriction on the power of our analysis, we no longer only keep a single matched placebo dying inventor for each real death. Instead, we keep all the matches, but drop duplicated placebo matches who match to multiple dying inventors. In Panel (a) we focus on the impact of immigrant death on native co-authors. Panel (b) shows the impact of immigrant deaths of native co-authors. Focusing on the percent change in output, we see that for all four outcomes an immigrant death has quite similar effects on both native co-authors and immigrant co-authors. However, our power here is a bit diminished. Turning to the deaths of natives. Panel (c) and Panel (d) show their impacts on native and immigrant collaborators, respectively. While our estimates are a bit noisy, it does not appear that the immigrant status of the surviving collaborator matters for the productivity loss due to the dying inventor.

Finally, we explore how inventor deaths affect the size of the team of her collaborators. If an inventor experiencing the death of a prior collaborator goes out and seeks a replacement, the net collaborator decline will be less than one. Alternatively, if losing a prior collaborator makes it harder to connect to new collaborators, the decline could be greater than one. To do so, in table 5 we repeat the inventor death analysis, but replace the dependant variable in these difference-in-difference regressions with a count of the number of living collaborators.

In panel (a) of Table 6 we explore the impact of an native inventor death on team sizes of collaborators. In column (1) we find that when a native loses a prior native collaborator, they lose 1.29 total natives and 0.06 immigrant co-authors (column 2), implying a total loss of additional 0.35 collaborators beyond the deceased native inventor. In contrast, column (3) illustrates that a death of a native leads co-authors to lose 0.94 natives and gain 0.02 immigrant co-authors (column 4), implying that these dying immigrants essentially are not replaced, but also do not lead to further losses in future collaborators.

5.4 Investigating Mechanisms

To explore potential mechanisms driving the differential productivity impacts of native versus immigrant inventor deaths on co-authors, we re-estimate the inventor death effects after re-weighting the immigrant and native deceased inventor samples to look more similar on observable characteristics.

One potential explanation for the heterogeneous effects of immigrants and natives is that US immigration policy cream skims the most productive people from the rest of the world and, therefore, the average immigrant inventor is simply of higher quality than the average native inventor. To the extent that more productive inventors have larger spillover effects, this could account for our results.

In order to test for this, we create a new sample of deceased natives that mirrors the productivity distribution of the deceased immigrants. We match deceased immigrants to deceased natives on ventiles of cumulative patents, deciles of cumulative citations, and the decade of age of death. All but three deceased immigrants have native matches based on these criteria. We drop these three immigrants from the analysis and randomly select up to 10 dying native matches per dying immigrant.[28] Table 7 shows the productivity of the matched deceased immigrant and native inventors up until the time of death in Panel A. In Panel B we find that the collaborators of this matched sample are also very similar in their pre-death productivity measures, and significantly more so than in the unmatched sample used in the above analysis.

We then repeat our analysis above with this new sample and find that the magnitude of the native effects only modestly increase when we are matching on the productivity distribution. We focus on number of patents and adjusted citations outcomes since these are the dimensions we matched the samples on. The results are reported in Table 8. The death of these higher ability natives has an only slightly higher impact on collaborators' number of patents as that of the average native death as reported in column (3) of Table 4, leading to an 11% decline instead of a 9% decline. Thus, differences in productivity between immigrants and natives explain very little of the the gap in collaboration externalities as measured by number of patents. When looking at adjusted citations, we see a 4 percentage point larger impact of native death, 8% in Table 4 vs 12% in our re-weighted sample, still statistically insignificant and much smaller than the immigrant effect of 32%.

We next investigate an alternative explanation: differences in collaborators' prior knowledge. As we saw in our descriptive analysis, immigrants' patents are more likely to cite foreign patents, indicating that immigrants likely have differing knowledge than the average native inventor. Thus, it is possible immigrants' collaboration externalities are quite large due to the fact that they have different knowledge backgrounds than their collaborators and there are significant synergies in knowledge sharing.

To measure the gap in knowledge backgrounds between a pair of collaborating inventors at the time of death, we make a list of all patents cited by each inventor across each of their own patents up until that point. We then compute the share of citations the two collaborators' prior patents have in common. If two collaborators work in the same sub-field, each of their own prior patents would likely list many of the same patents in their bibliographies. However, if two collaborators come from quite different fields of innovation, there will likely be little overlap between the collaborators' citations in their patents' bibliographies.

We now match immigrant deaths to native deaths according to the productivity metrics discussed above, as well as deciles of the constructed knowledge gap distribution. We then re-estimate the native death effects, which now match the immigrants in terms of both knowledge distance and productivity. The results are in Table 9. Since our matching metrics are based on citations, and number of patents, we focus on these outcomes as our main metrics of interest. We now find significantly smaller gaps between the treatment effects of immigrant and native deaths. In terms of number of patents, our baseline native estimate was 9%, which now increases to 16%. Using this metric, the knowledge gap differences between immigrants and native can explain 41% of the

[28]We keep up to ten native matches per immigrant to maximize the power of our analysis. All analysis weights the matched natives by one over the number of matched natives to each dying immigrant to keep the two treatment samples balanced on the match characteristics.

difference in the estimated collaboration externalities between these groups. The results are even more stark for adjusted citations. Our baseline results using adjusted citations found that the average native death lowered the productivity of collaborators by a statistically insignificant 9%, while the immigrant effect was 31%. Using our re-weighted sample, we find the matched native death lowers collaborators' adjusted citations by 36%. Thus, using either of these metrics, we conclude that including knowledge gap differences between immigrants and natives can explain much more of their heterogeneous collaboration externalities than productivity differences alone can.

6 Decomposition of Immigrant Contribution to US Innovation

The previous sections showed that immigrants have substantial contributions to US innovation, both directly through their own output and indirectly through positive spillovers onto their collaborators. To quantify the share of innovation which can be attributed to immigrants, we combine these estimates and conduct a back-of-the-envelope calculation using a simple framework.

6.1 Simple Framework

Suppose that the innovative output of an inventor i in year t is given by a Cobb-Douglas production technology:

$$Y_{it} \quad = \quad A_{it}\left(1+N_{i,nat}\right)^{\beta_{nat}}\left(1+N_{i,imm}\right)^{\beta_{imm}}. \tag{3}$$

Here, β_n for $, n \in \{nat, imm\}$ captures the (proportional) indirect productivity boost of a co-author of type n. The number of unique prior native and immigrant co-authors of inventor i in his career as of year t are given, respectively, by $N_{i,nat}$ and $N_{i,imm}$. A_{it} reflects the direct innovative productivity of inventor i in year t, as captured by inventor ability, training and prior experience.

We take a first order approximation of these production functions to map our estimates of co-author deaths to these production function parameters.[29] Taking the first order Taylor expansion around a base value and rearranging gives:

$$\frac{Y_{it}-\bar{Y}}{\bar{Y}} = \beta_{nat}\frac{\left(N_{i,nat}-\bar{N}_{nat}\right)}{\left(1+\bar{N}_{nat}\right)} + \beta_{imm}\frac{\left(N_{i,imm}-\bar{N}_{imm}\right)}{\left(1+\bar{N}_{imm}\right)} + \varepsilon_{it} \tag{4}$$

where all variables with a bar over them are measured at the point we are taking the first order Taylor expansion around. These equations offer intuition on how we should interpret the magnitude of our reduced form death estimates. The left-hand side of these equations represent the percent decline in output due to the change in the number of (living) collaborators. These are the exact number

[29]The reason we use a first order approximation is to simplify the issues of dealing with years when inventors have zero output. This prevents us from taking logs of the production functions. Working with the production function directly in levels would deliver a model where the error term A_{it} would be non-separable, making structural estimation challenging. Since we focus here more on a back-of-the envelope approach, the first order approximation makes it simple to make our OLS estimates to parameters of this production function.

estimated in our difference analysis, presented in Table 4. For example, we found that the death of an immigrant co-author lowered surviving collaborators' productivities by 32% in terms of adjusted citations. The right-hand side of these equations show that this productivity decline depends on the percent change in immigrant and non-immigrant co-authors, scaled by the production function parameters, β_m. This highlights that to estimate the production function parameters, we need to know how the exogenous death of a prior collaborator changes to total number of (living) prior immigrant and native collaborators. This is estimated in Table 6. For example, we found that the death of an immigrant collaborator leads to further lower the number of collaborators by 0.24 natives and 0.04 immigrants.

We next take these estimates of co-author loss and our estimated productivity losses and plug them into the first order approximation equations above to recover the parameters of the innovation production function. These estimates are in Table 10.[30] Consistent with our reduced form findings above, we find immigrant collaborators have much larger spillover effects onto their collaborators' productivities.

Next, we use our production function parameter estimates to back out the "ability" differences between immigrants and natives (their A_{it}s), by setting their co-authors to zero. Panel B of Table 8 reports the mean A_{it}s for immigrants and natives. As previously shown in Table 2, the average immigrant inventor produced about 0.16 more adjusted citations per year. However, once we strip out the spillover effects of collaborators and compare the average ability between immigrant and native inventors, we see immigrants would only produce 0.007 more adjusted citations per year than native inventors. Since immigrants have large co-author networks and more immigrant collaborators, they are able to even more productive than their "raw ability" advantage over natives would indicate.

6.2 Decomposing Aggregate Innovation

Finally, we use our model to decompose the channels through which immigrants and native contribute to total US innovation. We focus on adjusted citations as our metric. We want to highlight that these calculations are a accounting decomposition of the observed innovation we see in the data. These do not represent counterfactaul analysis of what would have happened had we not had immigrants in the US.

First, we quantify the importance of immigrants' indirect contribution to native production. As a starting point, we see that natives produce 78 percent of the total innovation in the data, as we showed in Figure 1. Next, we calculate how much native inventors would have produced had they had zero immigrant collaborators, holding fixed their number of native co-authors. Specifically, this implies that using the parameters of the production function estimated above, we calculate

[30] Specifically, we have two equations that are similar to equation (3) above, that highlight the effect of immigrant/native deaths on their collaborators. To calculate the parameters β_m, for $m \in \{nat, imm\}$, we use estimates from Tables 5 and 6. First, we use the estimates of the effects of a native death on adjusted citations from Table 5 to calculate $\frac{Y_{it}^{native} - \bar{Y}^{native}}{\bar{Y}^{native}}$. Second, we calculate the effects of native death on unique number of collaborators from Table 6 to estimate $N_{i,m} - \bar{N}_m$. We also use the average number of unique collaborators before death, $\bar{N}_m$. Ultimately, we solve two equations and two unknowns to extract the production function β_m parameters.

aggregate innovation of natives, assuming no immigrant collaborators. Specifically, we calculate:

$$\frac{\sum_{i \in \mathbb{I}_{nat}} A_{it} \left(1 + N_{it,nat}\right)^{\beta_{nat}}}{Y_{agg}} \tag{5}$$

When Y_{agg} is aggregate innovation and $\mathbb{I}_{nat}$ is the pool of native inventors. In column 2 of Table 11 shows that natives' innovation would now fall to 55% of total innovation. Thus, 23 percent (78%-55%) of total US innovation can be attributed to immigrants' collaboration spillovers on their native collaborators. Hence, this implies that 31% ((78-54)/78) of natives' total innovation can be indirectly attributed to their immigrant co-authors.

Second, we explore the importance of native's indirect contribution to immigrants' production. As shown in Table 11, immigrants produce 22 percent of total innovation in the data, in measured by observed output. Next, we calculate how much immigrant inventors would have produced had they had zero native collaborators, holding fixed their number of immigrant co-authors. Specifically, we calculate:

$$\frac{\sum_{i \in \mathbb{I}_{imm}} A_{it} \left(1 + N_{it,imm}\right)^{\beta_{imm}}}{Y_{agg}} \tag{6}$$

When Y_{agg} is aggregate innovation and $\mathbb{I}_{imm}$ is the pool of immigrant inventors. As reported in column (2) of Table 11, immigrants would have produced 15% of aggregate innovation, implying natives' spillovers onto immigrants' productivity accounts for 7% of aggregate innovation.

All in all, these calculations suggest that immigrants contribute directly to 15% of innovation, and their indirect contributions, through the enhancement of natives' productivity, explains 23% of innovation. All in all, immigrants account for 37% of total US innovation, despite only making up 16 percent of the inventor workforce. Further, the cross-group spillovers (natives on immigrants and immigrants on natives) account for 30 percent of total US innovation, highlighting the importance of combining these different immigrant and native knowledge bases to produce more aggregate innovation.

7 Conclusion

In this paper, we characterize the contribution of immigrants to the innovative output of the United States since 1976. Using inventor address information provided by the USPTO, we link patent records to the Infutor database first described in Diamond et al. (forthcoming). We then develop a novel methodology based on the first five digits of an individual's SSN and the individual's year of birth to identify the immigrant status of inventors. We perform several validation checks of this procedure and show that our methodology matches Census provided county immigrant shares with a very high degree of accuracy.

We find that over the course of their careers, immigrants are more productive than natives, as measured by number of patents, patent citations, and the economic value of these patents. Immigrant inventors also appear to facilitate the importation of foreign knowledge into United States, with immigrants inventors relying more heavily on foreign technologies and collaborating more with foreign inventors. Immigrant inventors have a greater number of collaborators than native inventors and while they are more likely to work with other immigrants, this tendency declines over time.

Using an identification strategy that exploits premature deaths, we show that immigrant inventors also contribute to the innovative productivity of the United States through their positive spillover effects on other US-based inventors. Our results highlight the importance of bringing high-skilled immigrants to the US to collaborate with US-born inventors, enabling them to combine their knowledge bases and push forward the innovation frontier.

References

Abramitzky, Ran and Leah Boustan, "Immigration in American economic history," *Journal of economic literature*, 2017, *55* (4), 1311–45.

_ , Leah Platt Boustan, and Katherine Eriksson, "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration," *American Economic Review*, 2012, *102* (5), 1832–56.

_ , _ , and _ , "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration," *Journal of Political Economy*, 2014, *122* (3), 467–506.

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

_ , John Van Reenen, and Luigi Zingales, "Innovation and institutional ownership," *American Economic Review*, 2013, *103* (1), 277–304.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas, "Immigration and the rise of american ingenuity," *American Economic Review*, 2017, *107* (5), 327–31.

Arrow, Kenneth, "Economic welfare and the allocation of resources for invention," in "The rate and direction of inventive activity: Economic and social factors," Princeton University Press, 1962, pp. 609–626.

Audretsch, David B and Maryann P Feldman, "R&D spillovers and the geography of innovation and production," *The American economic review*, 1996, *86* (3), 630–640.

Azoulay, Peter, Joshua S. G. Zivin, and Gustavo Manso, "Incentives and Creativity: Evidence from the Academic Life Sciences," *Working Paper*, 2010.

Balsmeier, Benjamin, Alireza Chavosh, Guan-Cheng Li, Gabe Fierro, Kevin Johnson, Aditya Kaulagi, Doug O'Reagan, Bill Yeh, and Lee Fleming, "Automated disambiguation of us patent grants and applications," *Working Paper*, 2015.

Basilio, Leilanie, Thomas K Bauer, and Anica Kramer, "Transferability of human capital and immigrant assimilation: An analysis for Germany," *Labour*, 2017, *31* (3), 245–264.

Becker, Gary S, "Human capital theory," *Columbia, New York*, 1964, *1964.*

Becker, Howard Saul, *Art worlds*, Univ of California Press, 1982.

Becker, Sascha O and Hans K Hvide, "Do entrepreneurs matter?," 2013.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "The lifecycle of inventors," 2016.

Ben-Porath, Yoram, "The production of human capital and the life cycle of earnings," *Journal of political economy*, 1967, *75* (4, Part 1), 352–365.

Bennedsen, Morten, Francisco Pérez-González, and Daniel Wolfenzon, "Do CEOs Matter?," 2008.

Bernstein, Shai, "Does going public affect innovation?," *The Journal of Finance*, August 2015, *70* (4), 1365–1403.

_ **, Emanuele Colonnelli, Davide Malacrino, and Tim McQuade**, "Who Creates New Firms when Local Opportunities Arise?," 2018.

Borjas, George J, *Immigration economics*, Harvard University Press, 2014.

Chellaraj, Gnanaraj, Keith E Maskus, and Aaditya Mattoo, "The contribution of skilled immigration and international graduate students to US innovation," 2005.

_ **,** _ **, and** _ , "The contribution of international graduate students to US innovation," *Review of International Economics*, 2008, *16*, 444 – 462.

Diamond, Rebecca, Timothy McQuade, and Franklin Qian, "The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco," *American Economic Review*, forthcoming.

Doran, Kirk Bennett, Alexander Gelber, and Adam Isen, *The Effect of High-Skilled Immigration on Patenting and Employment: Evidence from H-1B Visa Lotteries*, National Bureau of Economic Research, 2014.

Ellison, Glenn, Edward Glaeser, and William Kerr, "VWhat Causes Industry Agglomeration? Evidence from Coagglomeration PatternsV," *NBER Working Paper*, 2007, *13068*, 17.

Fadlon, Itzik and Torben Heien Nielsen, "Family Labor Supply Responses to Severe Health Shocks," Technical Report, National Bureau of Economic Research 2015.

Foley, C Fritz and William R Kerr, "Ethnic innovation and US multinational firm activity," *Management Science*, 2013, *59* (7), 1529–1544.

Grogger, Jeffrey and Gordon H Hanson, "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 2011, *95* (1), 42–57.

_ **and** _ , "Attracting talent: Location choices of foreign-born PhDs in the United States," *Journal of Labor Economics*, 2015, *33* (S1), S5–S38.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg, "The NBER patent citations data file: Lessons, insights and methodological tools," *Working Paper*, 2001.

_ **,** _ **, and** _ , "Market value and patent citations," *RAND Journal of Economics*, April 2005, *36* (1), 16–38.

Hunt, Jennifer, "Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa," *Journal of Labor Economics*, 2011, *29* (3), 417–457.

_ **and Marjolaine Gauthier-Loiselle**, "How much does immigration boost innovation?," *American Economic Journal: Macroeconomics*, 2010, *2* (2), 31–56.

Isen, Adam, "Dying to Know: Are Workers Paid Their Marginal Product?," *Unpublished manuscript*, 2013.

Jaffe, Adam B, "Real effects of academic research," *The American economic review*, 1989, pp. 957–970.

Jaffe, Adam B. and Manuel Trajtenberg, *Patents, citations, and innovations: A window on the knowledge economy*, Cambridge and London: MIT Press, 2002.

Jaravel, Xavier, Neviana Petkova, and Alex Bell, "Team-specific capital and innovation," *American Economic Review*, 2018, *108* (4-5), 1034–73.

Jones, Benjamin, EJ Reedy, and Bruce A Weinberg, "Age and scientific genius," Technical Report, National Bureau of Economic Research 2014.

_ , _ , and _ , "Age and scientific genius," Technical Report, National Bureau of Economic Research 2014.

Jones, Benjamin F, "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?," *The Review of Economic Studies*, 2009, *76* (1), 283–317.

_ , "Age and great invention," *The Review of Economics and Statistics*, 2010, *92* (1), 1–14.

Jones, Benjamin F. and Benjamin A. Olken, "Do Leaders Matter? National Leadership and Growth Since World War II*," *The Quarterly Journal of Economics*, 2005, *120* (3), 835–864.

Jones, Benjamin F and Bruce A Weinberg, "Age dynamics in scientific creativity," *Proceedings of the National Academy of Sciences*, 2011, p. 201102895.

Kerr, William R, "The ethnic composition of US inventors," 2008.

_ , "Ethnic scientific communities and international technology diffusion," *The Review of Economics and Statistics*, 2008, *90* (3), 518–537.

_ , "The agglomeration of US ethnic inventors," in "Agglomeration economics," University of Chicago Press, 2010, pp. 237–276.

_ **and William F Lincoln**, "The supply side of innovation: H-1B visa reforms and US ethnic invention," *Journal of Labor Economics*, 2010, *28* (3), 473–508.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Lanjouw, Jean O., Ariel Pakes, and Jonathan Putnam, "How to count patents and value intellectual property: The uses of patent renewal and application data," *Journal of Industrial Economics*, 1998, *46* (4), 405–432.

Lehman, HARVEY C, "Age and Achievement American Philosophical Society," 1953.

Lerner, Josh, Morten Sorensen, and Per Strömberg, "Private equity and long-run investment: The case of innovation," *Journal of Finance*, April 2011, *66* (2), 445–477.

Levin, Sharon G and Paula E Stephan, "Research productivity over the life cycle: Evidence for academic scientists," *The American Economic Review*, 1991, pp. 114–132.

Marshall, Alfred, *Some aspects of competition*, Harrison and Sons, 1890.

McDowell, John M, "Obsolescence of knowledge and career publication profiles: Some evidence of differences among fields in costs of interrupted careers," *The American Economic Review*, 1982, *72* (4), 752–768.

Moser, Petra, "Taste-based discrimination evidence from a shift in ethnic preferences after WWI," *Explorations in Economic History*, 2012, *49* (2), 167–188.

_ **, Alessandra Voena, and Fabian Waldinger**, "German Jewish émigrés and US invention," *American Economic Review*, 2014, *104* (10), 3222–55.

Nguyen, Bang Dang and Kasper Meisner Nielsen, "The value of independent directors: Evidence from sudden deaths," *Journal of Financial Economics*, 2010, *98* (3), 550–567.

Oettl, Alexander, "Reconceptualizing stars: Scientist helpfulness and peer performance," *Management Science*, 2012, *58* (6), 1122–1140.

Oster, Sharon M and Daniel S Hamermesh, "Aging and productivity among economists," *Review of Economics and Statistics*, 1998, *80* (1), 154–156.

Puckett, Carolyn, "The story of the social security number," *Soc. Sec. Bull.*, 2009, *69*, 55.

Romer, Paul M, "Endogenous technological change," *Journal of political Economy*, 1990, *98* (5, Part 2), S71–S102.

Saxenian, AnnaLee, "Silicon Valley's new immigrant high-growth entrepreneurs," *Economic development quarterly*, 2002, *16* (1), 20–31.

_ **, Yasuyuki Motoyama, and Xiaohong Quan**, *Local and global networks of immigrant professionals in Silicon Valley*, Public Policy Instit. of CA, 2002.

Seru, Amit, "Firm boundaries matter: Evidence from conglomerates and R&D activity," *Journal of Financial Economics*, February 2014, *111* (2), 381–405.

Simonton, Dean K, "Creative productivity through the adult years.," *Generations: Journal of the American Society on Aging*, 1991.

_ , "Emergence and realization of genius: The lives and works of 120 classical composers.," *Journal of Personality and Social Psychology*, 1991, *61* (5), 829.

Stephan, Paula and Sharon Levin, "Age and the Nobel Prize revisited," *Scientometrics*, 1993, *28* (3), 387–399.

Usher, Abbott Payson, "A history of mechanical innovations," *Cambridge, Mass*, 1954.

Weitzman, Martin L, "Recombinant growth," *The Quarterly Journal of Economics*, 1998, *113* (2), 331–360.

Zuckerman, Harriet, *Scientific elite: Nobel laureates in the United States*, Transaction Publishers, 1977.

Figure 1
Share of Immigrant Contribution

Categories are: (a) share in the overall population from 1990-2015 according to the ACS; (b) share of overall number of inventors, where inventor is defined as an individual who patent at least once; (c) share of overall number of patents; (d) share of overall number of citations, calculated over a three year horizon to avoid truncation issues; (e) citations normalized by the average number of citations in a given technology class year (the year in which all patents were applied); (f) share of top patents, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year; (g) share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only.

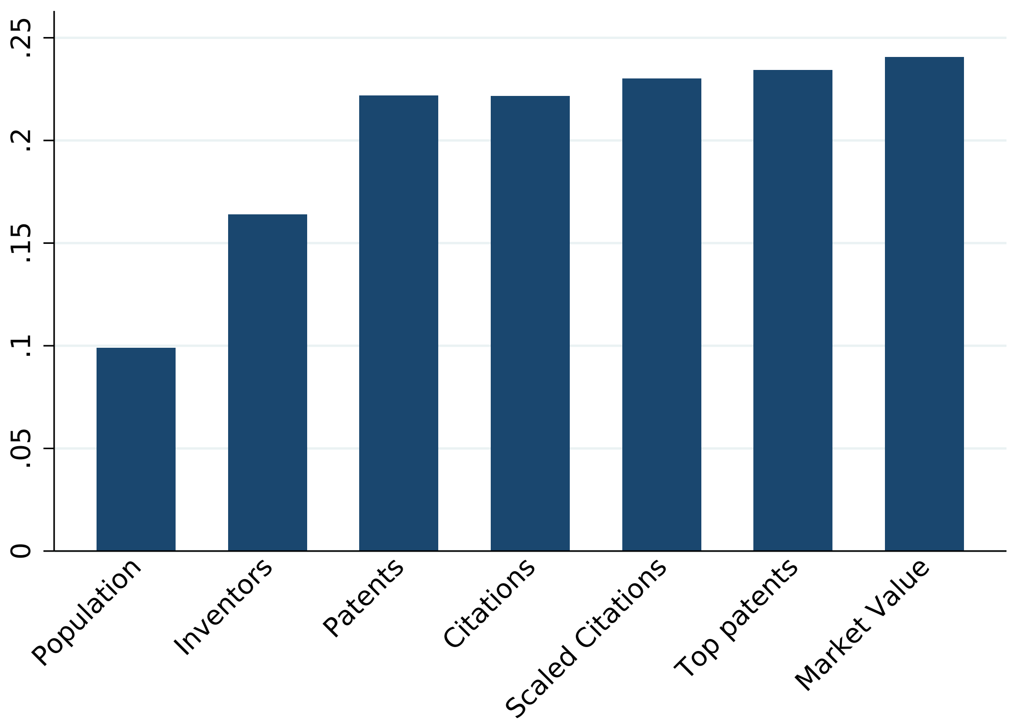
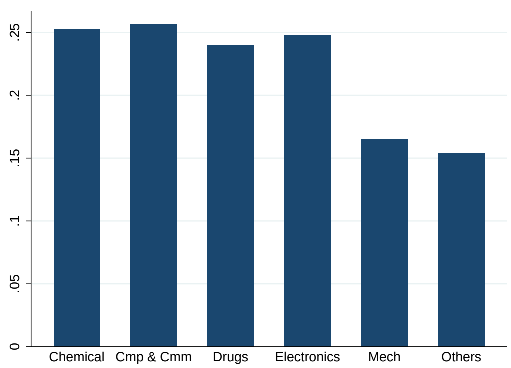


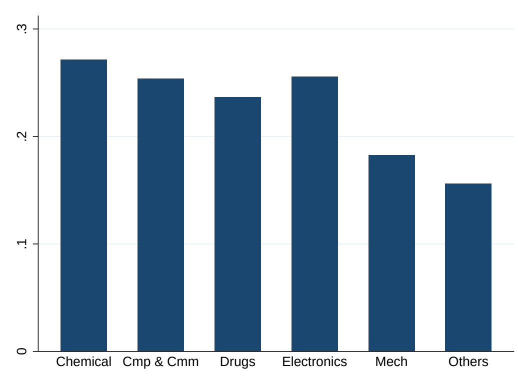
Figure 2
Share of Immigrant Contribution across Tech Classes

Categories are: (a) share of overall number of patents; (b) citations,calculated over a three year horizon to avoid truncation issues, normalized by the average number of citations in a given technology class year (the year in which all patents were applied); (c) share of top patents, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year; (d) share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only.

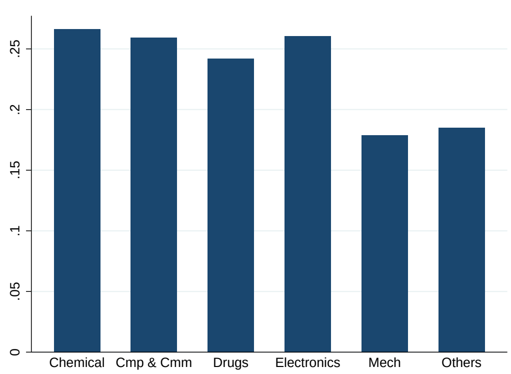
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

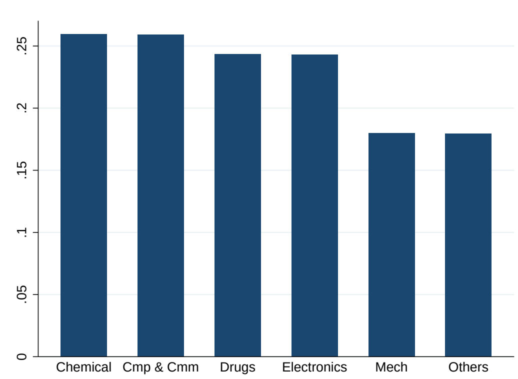
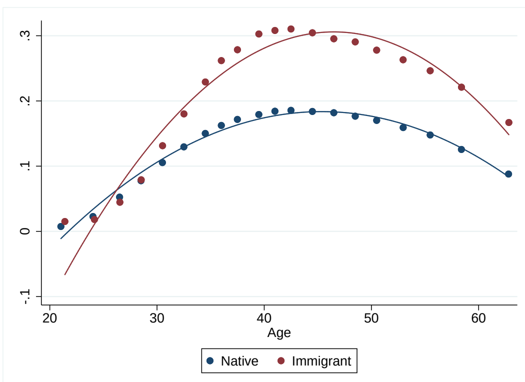


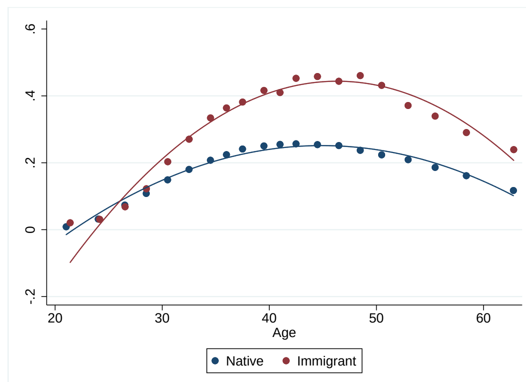
Figure 3
Productivity over the Life Cycle

Categories are: (a) share of overall number of patents; (b) citations,calculated over a three year horizon to avoid truncation issues, normalized by the average number of citations in a given technology class year (the year in which all patents were applied); (c) share of top patents, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year; (d) share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only.

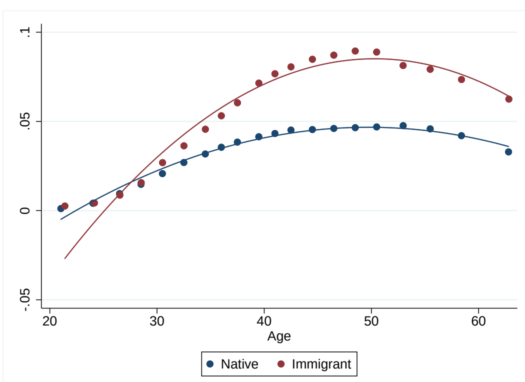
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

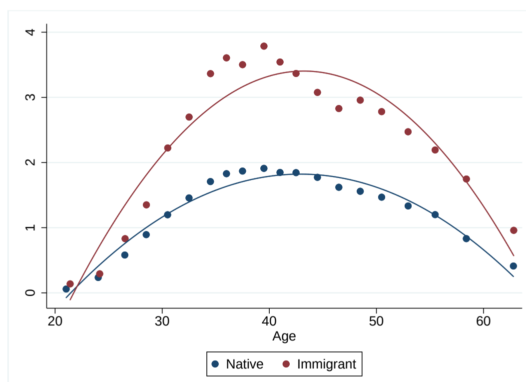
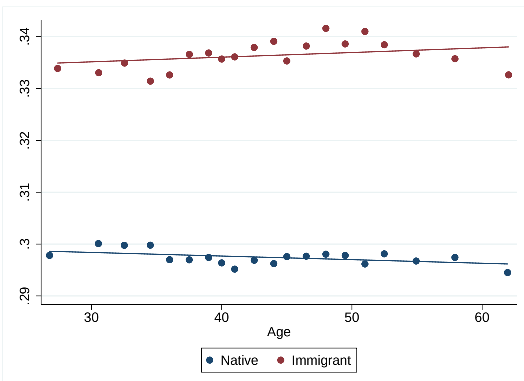


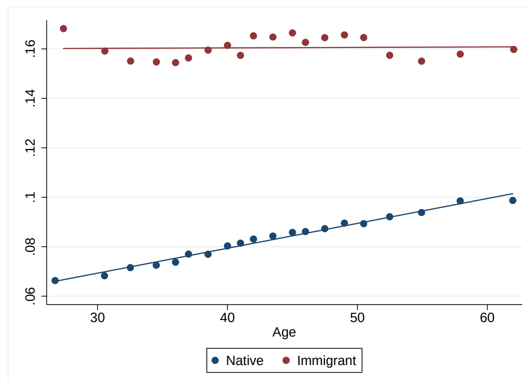
Figure 4
Global Knowledge Diffusion

Citations are calculated using a three year horizion to avoid trucation issues. Categories are: (a) share of foreign patents that were cited by the inventor in their patents; (b) share of patents in which a foreign inventor is one of the co-authors in a given year; (c) share of foreign patents that cited one of the inventors patents.

(a) **Share of Foreign Backward Citations**



(b) **Share of patents collaborated with foreign inventors**



(c) **Share of foreign forward citations**

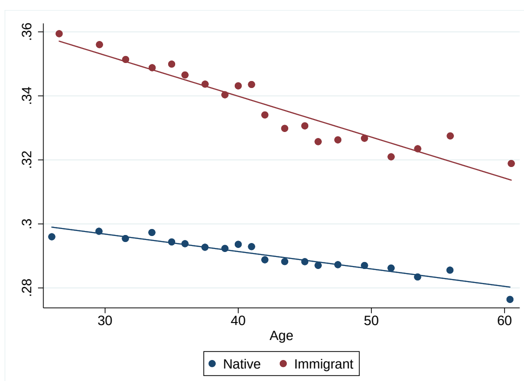
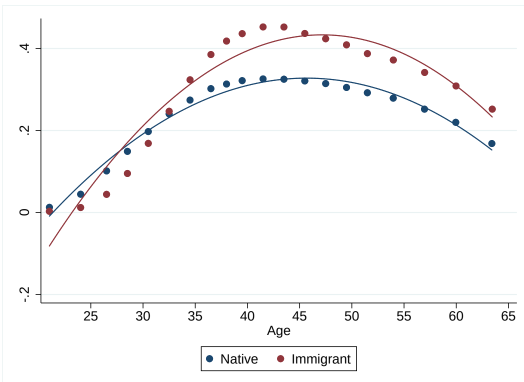


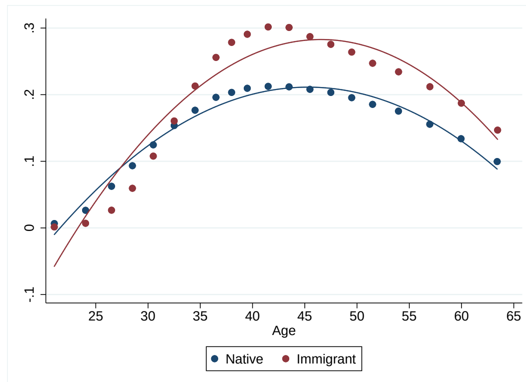
Figure 5
Assimilation over the Life Cycle

Categories are: (a) number of unique co-authors for all patents filled in a given year; (b) number of unique U.S. based co-authors for all patents filled in a given year (c) share of immigrants among unique co-authors for any given year ; (d) share of immigrants among unique U.S. based co-authors for any given year.

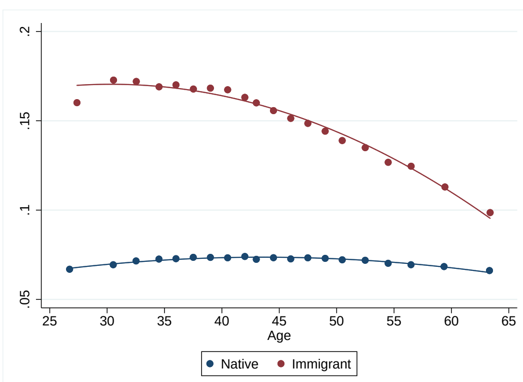
(a) **Number of unique co-authors**



(b) **Number of co-authors based in the U.S.**



(c) **Share of immigrant among unique co-authors**



(d) **Share of immigrant among unique U.S. based co-authors**

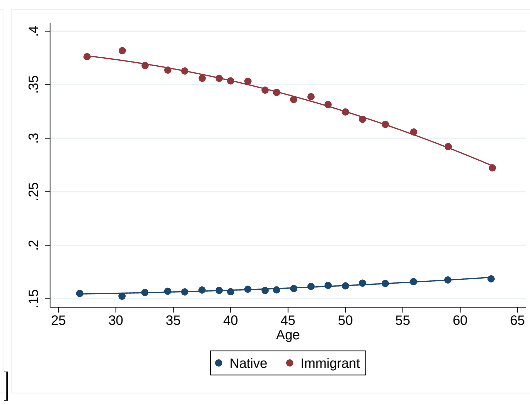
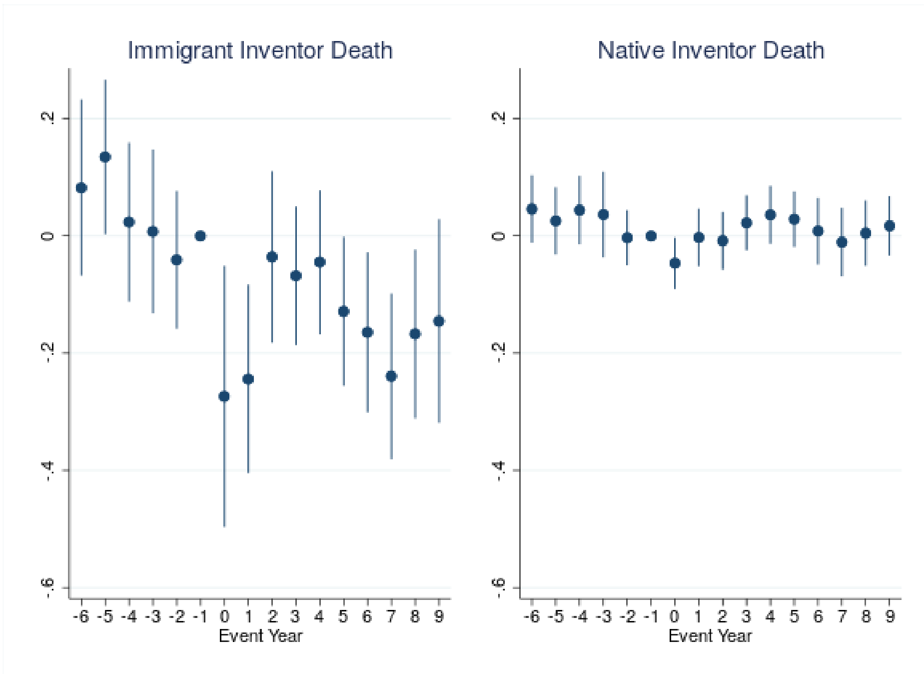


Figure 6
Comparing immigrant and native inventors death

Effect of the death of a co-author on inventor productivity for natives and immigrants, estimated using a diff-diff estimator in a sample matched by age, cumulative number of patents, year, ventiles of the number of co-authors. Vertical lines represent a 95% confidence interval constructed using standard errors clustered at the deceased inventor level. Categories are: (a) number of patents; (b) citations,calculated over a three year horizon to avoid truncation issues, normalized by the average number of citations in a given technology class year (the year in which all patents were applied).

(a) **Number of Patents**



(b) **Number of Adjusted Citations**

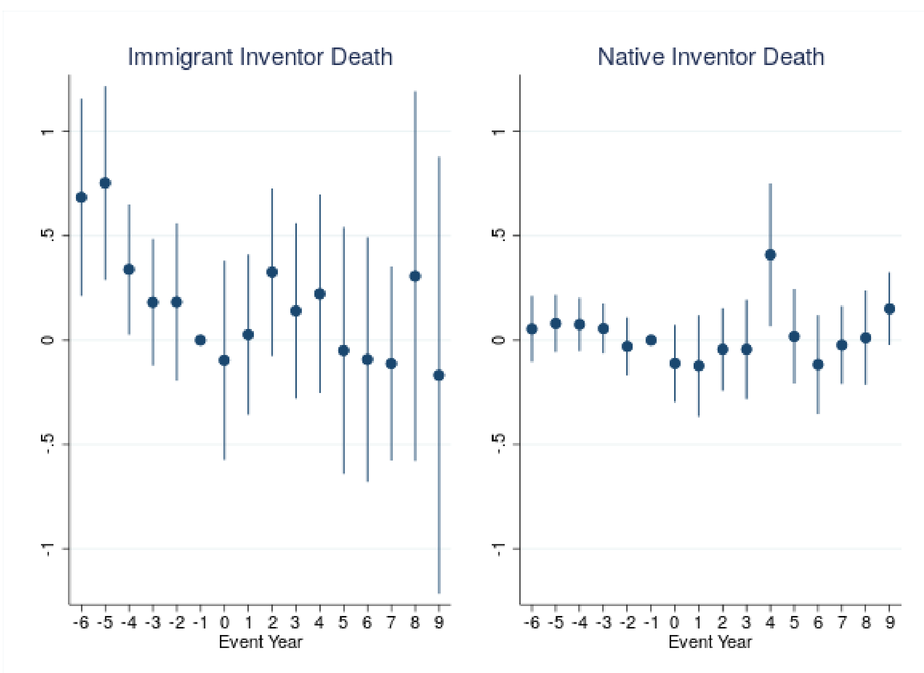
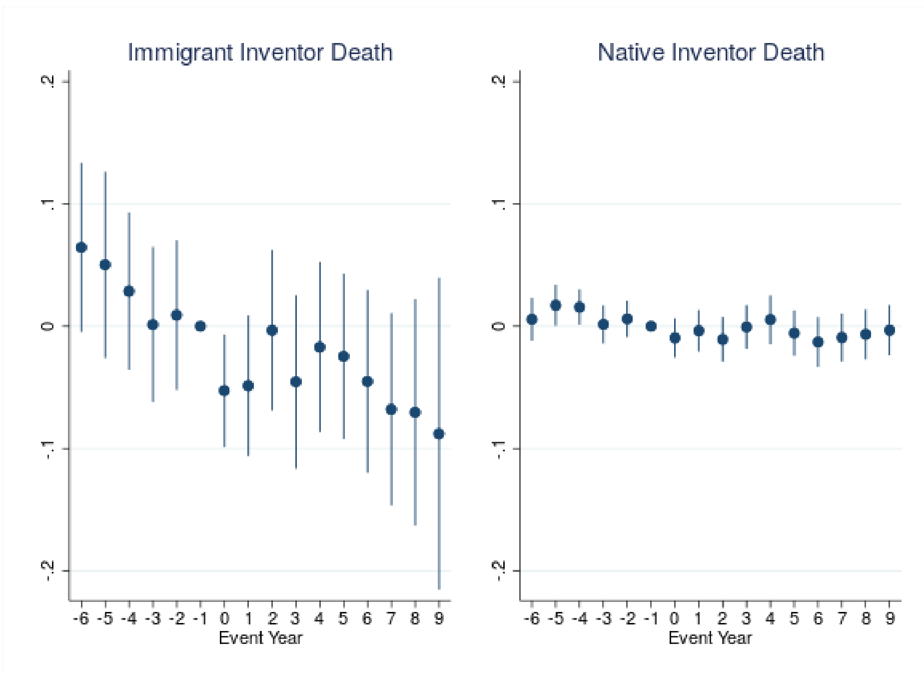


Figure 7
Comparing immigrant and native inventors death

Effect of the death of a co-author on inventor productivity for natives and immigrants, estimated using a diff-diff estimator in a sample matched by age, cumulative number of patents, year, ventiles of the number of co-authors. Vertical lines represent a 95% confidence interval constructed using standard errors clustered at the deceased inventor level. Categories are: (a) number of top patents, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year; (b) total patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only.

(a) **Top Patents**



(b) **Economic Value**

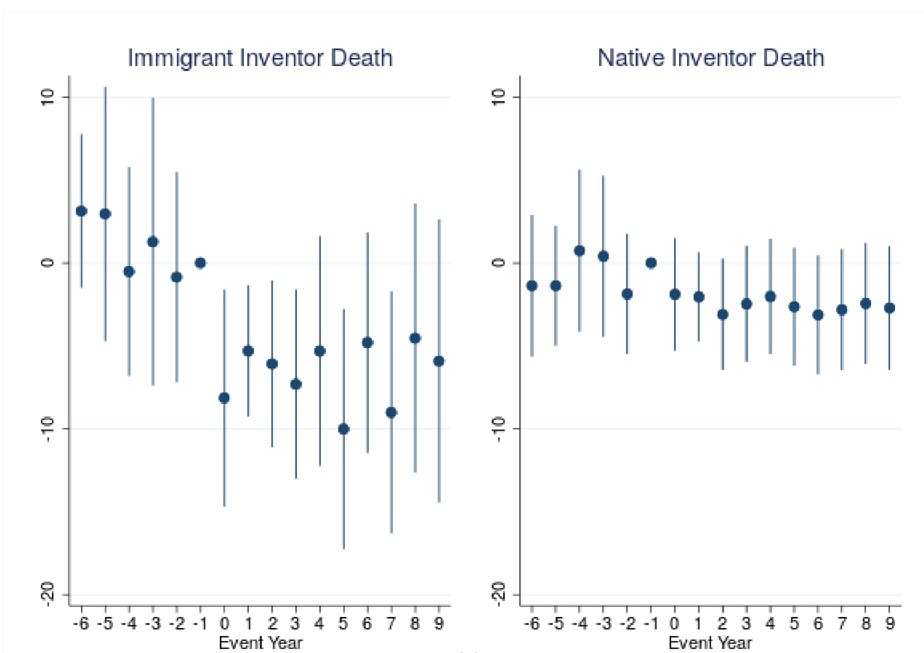


Table 1
Summary Statistics

This table shows summary statistics of the final inventor panel ranging from 1976 to 2012. *Number of Patents* is defined as the number of patents applied for by an inventor during the period. *Total Citations*is the total number of citations received by an inventor. *Total adjusted citations*is citations normalized by the average number of citations in a given technology class year (the year in which all patents were applied). *Total value created* is the share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only. *Top patents* is defined as a patent that is in the top 10% of citations in a given technology class and year. *Age at application* is the average age of all authors at the time of application.

Variables	Mean	Median	Top 90%	Std. Dev	# Obs.
Patenting Outcomes - Inventor-Level					
Number of patents	4.41	2.00	10.00	10.11	772876
Total citations	21.88	4.00	45.00	97.44	772876
Total adjusted citations	5.82	1.25	12.16	24.73	772876
Total value created	98.85	20.53	208.11	364.44	339348
Top patents	0.88	0.00	2.00	3.09	772876
Patenting Outcomes - Patent-Level					
Citations	4.47	2.00	11.00	9.97	1998644
Adjusted citations	1.22	0.52	2.86	3.66	1998643
Market value	18.42	7.20	38.48	49.61	910424
Top patents	0.19	0.00	1.00	0.39	1998644
Age at application	45.23	44.33	58.50	10.29	1998644
Demographics of Inventors					
Female	0.10	0.00	0.00	0.30	772876
Immigrant	0.16	0.00	1.00	0.37	772876

Table 2
Differential Sorting

This table estimates the effect of being an immigrant on inventors productivity with different combinations of fixed effects. Standard errors appear in parenthesis and are clustered at the inventor level. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively. Panel A shows the effect on total annual number of patents per-inventor. Panel B shows the effect on total annual citations normalized by the average number of citations in a given technology class year (the year in which all patents were applied). Panel C shows the effect on annual aggregate economic value of the patent, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only. Finally, panel D shows the effect on annual number of top patents, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year.

Panel A: Annual Number of Patents

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	0.106***	0.108***	0.0892***	0.0804***	0.0865***	0.0786***
	(0.00214)	(0.00214)	(0.00476)	(0.00423)	(0.00160)	(0.00159)
Observations	16,473,119	16,473,119	16,473,119	15,933,531	16,473,119	15,933,531
Year FE	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
County X Tech FE	no	no	no	yes	no	no
County X Year FE	no	no	no	no	yes	yes
Tech X Year FE	no	no	no	no	no	yes

Panel B: Annual Citation adjusted weighted Number of Patents

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	0.165***	0.168***	0.126***	0.114***	0.121***	0.109***
	(0.00675)	(0.00676)	(0.0104)	(0.00872)	(0.00453)	(0.00462)
Observations	16,473,119	16,473,119	16,473,119	15,933,531	16,473,119	15,933,531
Year FE	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
CountyXTech FE	no	no	no	yes	no	no
CountyXYear FE	no	no	no	no	yes	yes
Tech X Year FE	no	no	no	no	no	yes

Table 2
(Continued)

Panel C: Annual Aggregate Economic Value

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	1.416***	1.428***	1.083***	0.873***	1.074***	0.913***
	(0.0497)	(0.0497)	(0.107)	(0.0772)	(0.0476)	(0.0459)
Observations	16,473,119	16,473,119	16,473,119	15,933,531	16,473,119	15,933,531
Year FE	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
CountyXTech FE	no	no	no	yes	no	no
CountyXYear FE	no	no	no	no	yes	yes
Tech X Year FE	no	no	no	no	no	yes

Panel D: Annual Number of Top Patents

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	0.0312***	0.0323***	0.0253***	0.0236***	0.0242***	0.0226***
	(0.0008)	(0.0008)	(0.0018)	(0.0015)	(0.0007)	(0.0008)
Observations	16,473,119	16,473,119	16,473,119	15,933,531	16,473,119	15,933,531
Year FE	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
CountyXTech FE	no	no	no	yes	no	no
CountyXYear FE	no	no	no	no	yes	yes
Tech X Year FE	no	no	no	no	no	yes

Table 3
Inventor Death Controls

This table shows summary statistics for control variables and pre-treatment dependent variables for the real and placebo deceased and survivor inventors. The real and placebo deceased sample was created by matching on age, cumulative number of patents, year, and ventiles of the number of co-authors. In Panel A, controls include age, year of death, immigrant status, gender, team size, and number of teams. In Panel B, controls include age, immigrant status, and gender for the Infutor matched sample where the characteristics are available. For the full sample in Panel B, we also include collaboration strength variables: the number co-patents between a survivor inventor and his or her deceased co-inventor before time of death. Pre-treatment dependent variables in Panel A and Panel B are the same as described in Figure 7. Panel C shows the number of patents and share of patents for real and placebo deceased and survivor inventors in each of the six technology categories.

Panel A: Real vs Placebo Deceased Demographics

	Real Deceased			Placebo Deceased		
Variables	Mean	Median	Std. Dev	Mean	Median	Std. Dev
Age	51.5	53	7.3	51.5	53	7.3
Year	2001	2002	0.1	2001	2002	0.1
Immigrant status	0.09	0	0.29	0.09	0	0.29
Cumulative patents	2.8	1	6.1	2.8	1	5.6
Co-authors	5.7	1	18.9	5.5	1	14.9
Team size	1.9	1	1.5	1.9	1	1.5
Total adjusted citations	3.3	1	8.8	3.8	1	15.6
Top patents	0.42	0	1.3	0.48	0	1.9
Econ value	27.8	0	205.7	27.2	0	175.8
Number of teams	1.49	1	2.04	1.48	1	2.02
Female	0.07	0	0.3	0.09	0	0.3
Sample Size	9405			9405		

Table 3
(Continued)

Panel B: Real vs Placebo Co-Inventor Characteristics

	Real Deceased			Placebo Deceased		
Variables	Mean	Median	Std. Dev	Mean	Median	Std. Dev
Infutor Sample						
Age	51.3	50	11.7	49.3	49	11.0
Immigrant status	0.164	0	0.4	0.166	0	0.4
Female	0.10	0	0.3	0.09	0	0.28
Sample Size	16829			17477		
Full Sample						
Number of copatents pre-treat.	2.0	1	2.8	2.1	1	3.2
Cumulative patents	10.6	4	24.7	10.0	4	20.1
Total adjusted citations	15.0	4.25	39.5	15.8	4.4	56.5
Top patents	2.1	1	5.6	2.1	1	6.1
Econ value	158.2	12.6	730	140.0	11.8	504
Sample Size	27796			27393		

Panel C: Comparing Technologies

	Deceased Inventor		Placebo Inventor		Deceased Co-inventor		Placebo Co-inventor	
Tech Class	# Patents	Share	# Patents	Share	# Patents	Share	# Patents	Share
Chemicals	5150	19.5	4373	16.7	69918	23.3	63490	20.0
Computers	4883	18.5	5535	21.2	64936	21.6	80988	25.7
Drugs	3008	11.4	3745	14.3	42933	14.3	54647	17.3
Electronics	3752	14.2	4048	15.5	50867	16.9	55236	17.6
Mechanicals	4511	17.1	3766	14.4	36255	12.1	31323	9.9
Others	5126	19.4	4642	17.8	35551	11.8	30026	9.5

Table 4
Inventor Death

This table shows the diff-diff estimates of the inventor death full sample. The sample is the same as defined in table 3 and all variables are as defined in table 2. Standard errors appear in parentheses and are clustered at the deceased inventor level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Inventor Deaths

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Number of Patents	Number of Patents	Number of Patents	Top Patents	Top Patents	Top Patents
	All	Immigrants	Natives	All	Immigrants	Natives
Post Death	-0.0595***	-0.195***	-0.0420**	-0.0220***	-0.0708***	-0.0155**
	(0.0172)	(0.0500)	(0.0183)	(0.00586)	(0.0186)	(0.00615)
Control Post Mean	0.50	0.75	0.47	0.15	0.26	0.14
Percent Change	-12%	-26%	-09%	-15%	-27%	-11%
Observations	978797	120254	858543	978797	120254	858543
Match Group-Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes

Panel B: Inventor Deaths

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Adjusted Citations	Adjusted Citations	Adjusted Citations	Econ Value	Econ Value	Econ Value
	All	Immigrants	Natives	All	Immigrants	Natives
Post Death	-0.0912	-0.354**	-0.0557	-2.225***	-7.795***	-1.461*
	(0.0777)	(0.155)	(0.0859)	(0.712)	(2.058)	(0.757)
Control Post Mean	0.73	1.11	0.68	6.74	12.05	6.01
Percent Change	-12%	-32%	-08%	-33%	-65%	-24%
Observations	978797	120254	858543	978797	120254	858543
Match Group-Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes

Table 5
Inventor Death: Impact on Native and Immigrant Co-authors

This table shows the diff-diff estimates of the inventor death sample, breaking the effect into 4 categories: (a) the effect of a immigrant death on their native co-authors; and (b) the effect of a native death on their native co-authors. The sample and all variables are as defined in table 4. Standard errors appear in parentheses and are clustered at the deceased inventor level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Immigrant Death, Native Coauthors

	(1)	(2)	(3)	(4)
Dependent Variable	Number of Patents	Adjusted Citations	Top Patents	Econ Value
Post Death	-0.110*	-0.256*	-0.0502**	-3.646
	(0.0580)	(0.134)	(0.0219)	(2.361)
Control Post Mean	0.38	0.53	0.20	5.65
Percent Change	29%	48%	25%	64%
Observations	633757	633757	633757	633757
Match Group-Event Year FE	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes

Panel B: Immigrant Death, Immigrant Coauthors

	(1)	(2)	(3)	(4)
Dependent Variable	Number of Patents	Adjusted Citations	Top Patents	Econ Value
Post Death	-0.156	-0.481	-0.0629	-1.109
	(0.0994)	(0.304)	(0.0385)	(2.837)
Control Post Mean	0.59	1.02	0.32	11.28
Percent Change	27%	47%	19%	10%
Observations	286556	286556	286556	286556
Match Group-Event Year FE	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes

Table 5
(Continued)

Panel C: Native Death, Native Coauthors

	(1)	(2)	(3)	(4)
Dependent Variable	Number of Patents	Adjusted Citations	Top Patents	Econ Value
Post Death	-0.0306**	-0.0106	-0.00583	-1.120
	(0.0142)	(0.0398)	(0.00432)	(0.690)
Control Post Mean	0.39	0.49	0.13	4.73
Percent Change	8%	2%	8%	23%
Observations	7407542	7407542	7407542	7407542
Match Group-Event Year FE	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes

Panel D: Native Death, Immigrant Coauthors

	(1)	(2)	(3)	(4)
Dependent Variable	Number of Patents	Adjusted Citations	Top Patents	Econ Value
Post Death	-0.102*	-0.115	-0.0217*	-3.532*
	(0.0534)	(0.0736)	(0.0117)	(2.044)
Control Post Mean	0.68	0.98	0.23	10.66
Percent Change	15%	12%	10%	33%
Observations	1244056	1244056	1244056	1244056
Match Group-Event Year FE	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes

Table 6
Inventor Death: Impact on Number of Unique Collaborators

This table shows the diff-diff estimates of the inventor death sample, breaking the effect into 4 categories: (a) the effect of a immigrant death on their native co-authors; and (b) the effect of a native death on their native co-authors. The sample and all variables are as defined in table 4. Standard errors appear in parentheses and are clustered at the deceased inventor level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Dying Inventor	Native	Native	Immigrant	Immigrant
Prior Coauthors:	Native	Immigrant	Native	Immigrant
Post Death	-1.242***	-0.0401	-0.0345	-0.958***
	(0.0672)	(0.0270)	(0.207)	(0.0986)
Control Post Mean	7.85	1.45	6.16	3.37
Percent Change	16%	3%	1%	36%
Observations	517778	517778	69529	69529
Match Group-Event Year FE	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes

Table 7
Matching Deceased Natives and Deceased Immigrants

This table shows summary statistics for control variables and pre-treatment dependent variables for the matched deceased immigrant and native inventors. In Panel A, controls include age, year of death, immigrant status, gender, team size, and number of teams. In Panel B, controls include age, immigrant status, and gender for the Infutor matched sample where the characteristics are available. For the full sample in Panel B, we also include collaboration strength variables: the number co-patents between a survivor inventor and his or her deceased co-inventor before time of death. Pre-treatment dependent variables in Panel A and Panel B are the same as described in Figure 7. Panel C shows the number of patents and share of patents for real and placebo deceased and survivor inventors in each of the six technology categories.

Panel A: Native vs Immigrant Deceased Demographics

	Deceased Natives			Deceased Immigrants		
Variables	Mean	Median	Std. Dev	Mean	Median	Std. Dev
Age	52.3	54	6.8	52.6	54	6.8
Year	2001	2002	7.1	2000	2001	7.3
Immigrant status	0.0	0.0	0.0	1	1	0.0
Cumulative patents	3.76	2	7.2	4.2	2	8.5
Co-authors	5.4	3	7.3	5.3	3	7.5
Team size	1.9	1	1.5	1.8	1	1.3
Total adjusted citations	4.6	1.4	10.9	4.9	1.3	13.0
Top patents	0.61	0	1.7	0.67	0	1.9
Econ value	39.3	0	235	59.0	0	331
Number of teams	2.8	1	3.7	2.9	1	4.2
Female	0.06	0	0.24	0.16	0	0.36
Sample Size	5262			876		

Panel B: Deceased Natives vs. Deceased Immigrants Co-Inventor Characteristics

	Deceased Natives Co-Inventors			Deceased Immigrants Co-Inventors		
Variables	Mean	Median	Std. Dev	Mean	Median	Std. Dev
Infutor Sample						
Age	46.2	45	11.3	46.5	46	11.38
Immigrant status	0.15	0	0.4	0.31	0	0.5
Female	0.09	0	0.3	0.13	0	0.3
Sample Size	11397			2089		
Full Sample						
Number of copatents pre-treat.	2.1	1	3.0	2.2	1	3.0
Cumulative patents	13.4	4	48.0	13.8	5	29.0
Total adjusted citations	18.6	5.1	69.6	20.1	5.6	53.0
Top patents	2.5	1	9.1	2.9	1	7.84
Econ value	167	36.8	578	227	50.1	596
Sample Size	17195			3278		

Table 8

Productivity Matched Deceased Natives and Immigrants - all patents

This table shows the diff-diff estimates of the full sample when deceased immigrants are matched to deceased natives that have the same number of cumulative patents at the time of death. Standard errors appear in parentheses and are clustered at the deceased inventor level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Dependent Variable	Number of Patents	Number of Patents	Adjusted Citations	Adjusted Citations
	Immigrants	Natives	Immigrants	Natives
Post Death	-0.194***	-0.055**	-0.357**	-0.094
	(0.050)	(0.027)	(0.157)	(0.125)
Control Post Mean	0.75	0.53	1.11	0.78
Percent Change	-26%	-11%	-32%	-12%
Observations	119276	617755	119276	617755
Match Group-Event Year FE	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes

Table 9

Matched Deceased Natives and Immigrants - Knowledge Gap

This table shows the diff-diff estimates of the inventor death sample when deceased immigrants are matched to deceased natives that have the a similar number of cumulative patents and cumulative citations at the time of death, as well as a similar average knowledge gap with their coauthors. The sample is the same as defined in table 7 and all variables are as defined in table 2. Standard errors appear in parentheses and are clustered at the deceased inventor level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	(1) Number of Patents	(2) Number of Patents	(3) Adjusted Citations	(4) Adjusted Citations
	Immigrants	Natives	Immigrants	Natives
Post Death	-0.201***	-0.0818**	-0.341**	-0.379*
	(0.0515)	(0.0332)	(0.160)	(0.229)
Control Post Mean	0.75	0.52	1.11	1.03
Percent Change	-26%	-16%	-31%	-36%
Match Group-Event Year FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Observations	114797	496068	114797	496068

Table 10

Innovation Production Function Estimates

This table shows Cobb-Douglas innovation production function parameters estimates. These estimates come from our reduced form estimates of inventor deaths on collaborator productivity, depending on whether the dying inventor was an immigrant. These estimates use the imapcts on number of patents as the metric of interest.

Panel A: Production Function Parameter Estimates

	(1)	(2)
	β_{imm}	β_{nat}
	1.45	0.40
Inventor Type	Immigrant	Native

Panel B: Mean Ability Estimates (A_{it})

	(1)	(2)	(3)
	Immigrants	Natives	Difference
Avg A_{it}	0.037	0.030	0.007***
	(0.0003)	(0.0001)	(0.0004)

Table 11

Decomposing Aggregate Innovation Output

This table shows the direct and indirect contribution of natives and immigrants to total US innovation from 1976-2012. Estimates are based production function parameters reported in table above. Innovation is measured in terms of adjusted citations. Column 1 reports the observed output shares between immigrants and native in the data. Column 2 calculated output if immigrants only co-author with immigrants and natives only work with natives. Column 3 attributes the indirect effects of natives and immigrants on each other to those who are causing the increased output. Thus, Column 3 for immigrants equals immigrant output in Column 2 plus the change betwen columns 1 and 2, representing the additional output natives produce by working with immigrants.

	(1)	(2)	(3)
Native Output	0.78	0.55	0.63
Immigrant Output	0.22	0.15	0.37
Total Output	1.00	0.70	1.00
Direct Output Attribution:	YES	YES	NO
Indirect Output Attribution:	NO	NO	YES
Natives collaborate with:	Both	Natives	Both
Immigrants collaborate with:	Both	Immigrants	Both

Appendix

A Matching Algorithm of Patent Data with Infutor

The raw patent data obtained from Balsmeier et al. (2015) which links inventors over time. Similarly, the Infutor database links individuals over time, using Social Security numbers and other identifiers. We first subset the patent data to inventors who reside in the US. Before beginning the matching process, we first standardize the names in the patent data and Infutor. First, we split the "first name" column because originally the first and middle names are both saved in the "first name" column. Identifying the first space in the name, we split that string to first and middle names, and save the first part of the string as the clean first name and save the second part of the string (if it exists) as the clean middle name. Second, the suffixes "JR", "SR" and numerals "II", "III", "IV" often appear at the end of first and last names and are stripped out. Finally, we standardize the city names in both the patent data and Infutor by finding the preferred city name(s) from the US Postal Office for each city and state appeared in the data. Whenever the preferred city name is not available, we use the original city and state for matching.

Our matching algorithm is similar to Bell et al. (2016), we apply multiple steps to identify matches between inventors in the patent data and individuals in Infutor. In each step, inventors enter a match round only if they have not already been matched to an Infutor panelist in an earlier round. The share of data matched in each round is documented below.

- **Step 1**: Exact match on last name, state, city and the first three letters of the first name - 50.4% of inventors are uniquely one-to-one matched in this step. For these one-to-one matches, 94% have the same full first name. For those 6% where there is not an exact match on full first name, most of them seem to be cases where nicknames are used e.g. Fredrick vs Fred. This step also produces many-to-one matches in which multiple individuals in the Infutor data are matched to a single inventor. We disambiguate these many-to-one matches using several different ways in the next steps.

- **Step 2:** Exact match on last name, first name, state, and city - this step resolve some of the many-to-one matches originated in previous step, increasing the overall unique match rate to 55.7% of inventors in the patent database.

- **Step 3**: Exact match on last name, first three letters of the first name, middle name initial, state, and city - this step further disambiguates many-to-one matches, leading to an overall 64.8% match rate of inventors.

- **Step 4:** Exact match on last name, first name, middle name initial, state, and city - in this step we rely on both the full first name and the first letter of middle name, corresponding to a match rate of 65.5% of inventors.

- **Step 5**: Exact match on last name, first three letters of the first name, state, and city, as well as an overlap of the timing in which the address in Infutor is consistent with the address provided by Infutor - specifically, we require the patent application date to be no more than 180 days earlier than the first day of the recorded beginning month of residence for an Infutor address and no more than 180 days later than the first day of the recorded ending month of residence for an Infutor address. Adding this criterion brings the total match rate to 68.1% of inventors to be matched.

- **Step 6:** Exact match on last name, first two letters of the first name, state, and city for hitherto unmatched inventors - this enables additional matches and increasing the match rate to 68.2% of inventors to be matched.

- **Step 7:** Exact match on last name, first letter of the first name, state, and city for hitherto unmatched inventors - this step brings us to a total of 914,275 one-to-one matches of unique inventors that, corresponding to 68.5% of inventors to be matched.

Overall, our final match rate is 68.5% of inventors with a final sample of 914,725 unique inventors. As a comparison, Bell et al. (2016) match the patent database to federal income tax records, and obtain a match rate of 80% in the 1990s.

B Figures

Figure A.1
Age of immigrants at entry

Age of immigrants at the time of entry (based on receiving SSN)

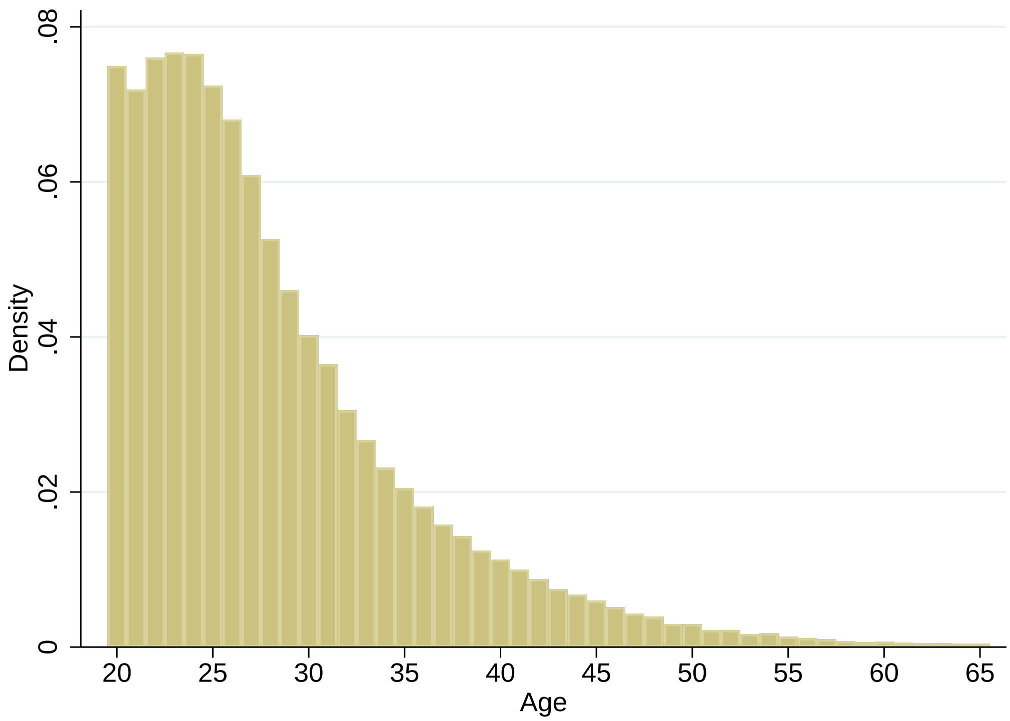


Figure A.2

Validation with CENSUS 2000 (only individuals older than 18) - Population Sizes (Millions)

Scatterplot at the County level. The y axis has the total population that is older than 18 years old in each County, according to the CENSUS 2000. The x axis has the number of people that *Infutor* placed living in each County in 2000. If *Infutor* places a person in two different Counties we use only the county in which that person stayed longer in 2000.

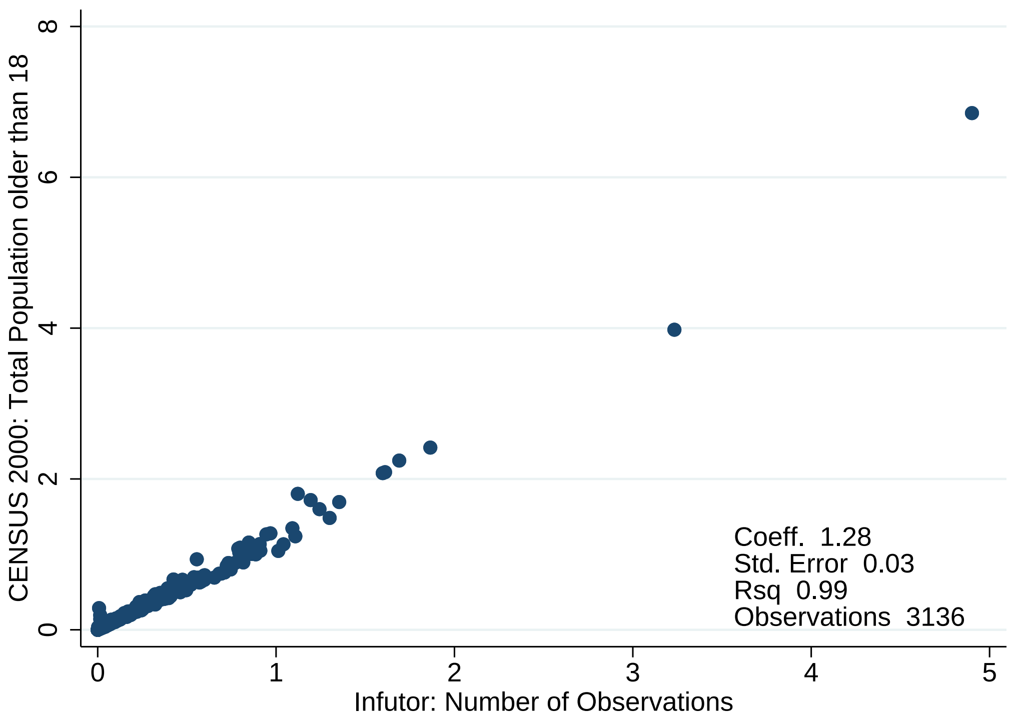


Figure A.3
SSN issuance age distribution

Quantiles of the age of SSN issuance distribution by assignment year, calculated at the individual level. Assignment year was colected from the website (https://www.ssn-verify.com/) for after 1950 and using the most frequent birth year plus 16 for before 1950. Data comes from *Infutor* only individuals that have a social security number and year of birth.

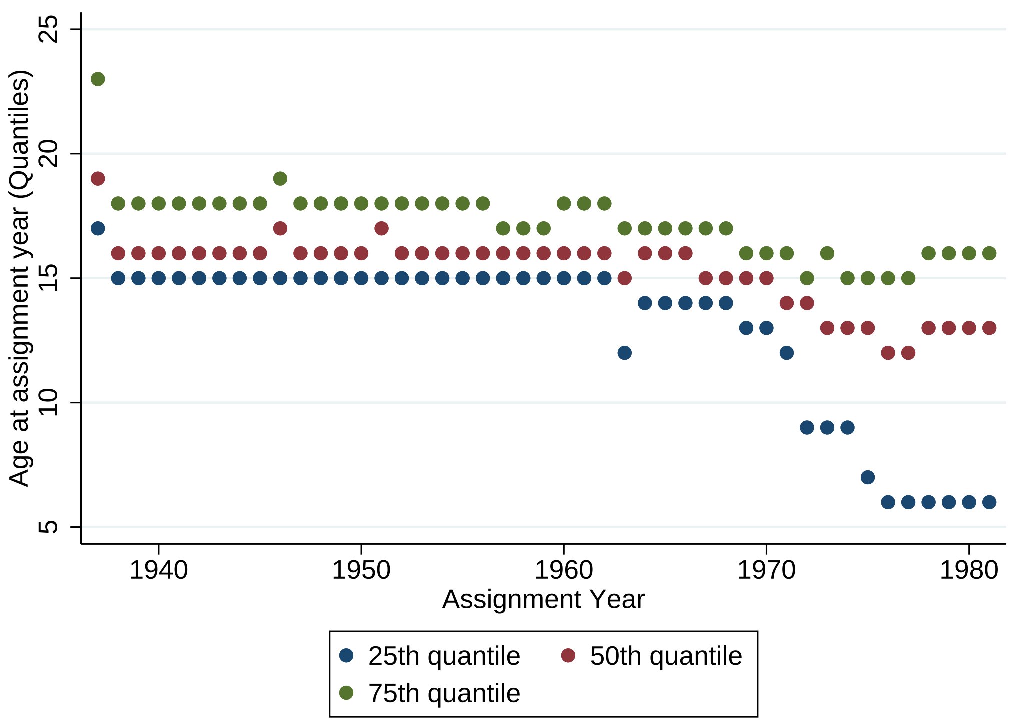


Figure A.4
Validation of the pre-1950 assignment year inputation

Binscatter of the encoded group numbers for each assignment year, constructed after controling for fixed effects of area code and weighted by the number of observations in each area and group. Assignment year was colected from the website (https://www.ssn-verify.com/) for after 1950 and using the most frequent birth year plus 16 for before 1950. Data comes from *Infutor* only individuals that have a social security number and year of birth.

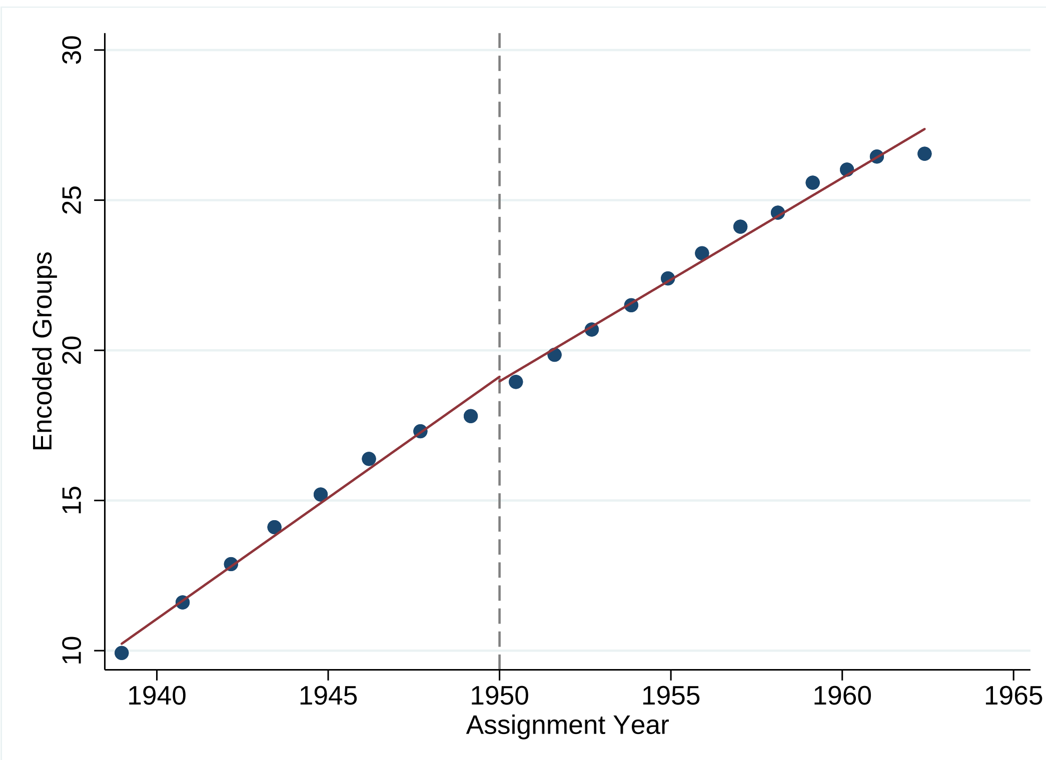


Figure A.5
Validation with the CENSUS 2000

R^2 and slope coefficient of regressing at the County level the proportion of foreign born in the CENSUS 2000 against the proportion of immigrants among all individuals that Infutor placed in County for each immigrant classification variable. The x-axis shows the minimum gap between assignment year and birth year needed to classify someone as immigrant for each immigrant classification variable. Data comes from *Infutor*, only individuals with a SSN number and a birth year. All regressions were weighted by the total population at that county at the CENSUS 2000.

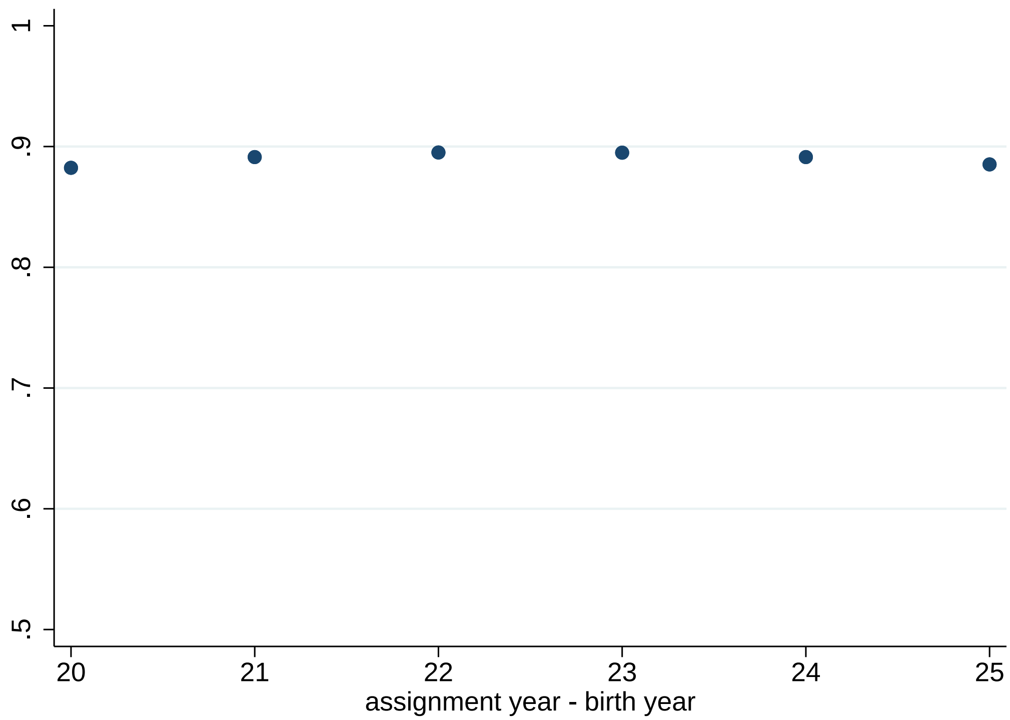
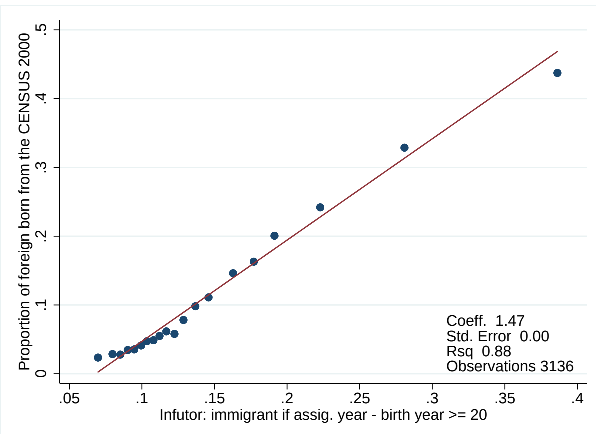


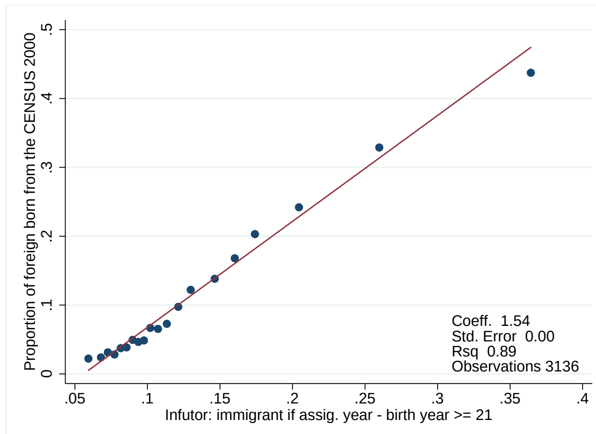
Figure A.6
Validation with the CENSUS 2000 - Binscatters

Binscatters of the proportion of foreign born in the CENSUS 2000 against the proportion of immigrants among all individuals that Infutor placed in County for selected immigrant classification variables at the county level. Data comes from *Infutor*, only individuals with a SSN number and a birth year. All regressions were weighted by the total population at that county at the CENSUS 2000.

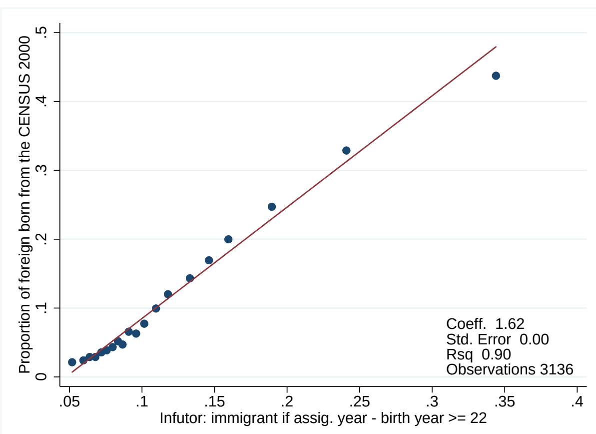
(a) **Immigrant if assig. year - birth year $\geq$ 20**



(b) **Immigrant if assig. year - birth year $\geq$ 21**



(c) **Immigrant if assig. year - birth year $\geq$ 22**



(d) **Immigrant if assig. year - birth year $\geq$ 23**

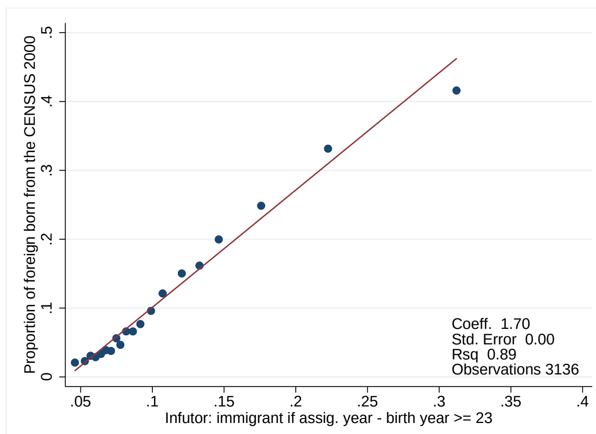


Figure A.7
Validation with the CENSUS 2000

R^2 and slope coefficient of regressing at the County level the proportion of foreign born in the CENSUS 2000 against the proportion of immigrants among all individuals that Infutor placed in County for each immigrant classification variable. The x-axis shows the minimum gap between assignment year and birth year needed to classify someone as immigrant for each immigrant classification variable. Data comes from *Infutor*, only individuals with a SSN number and a birth year. All regressions were weighted by the total population at that county at the CENSUS 2000.

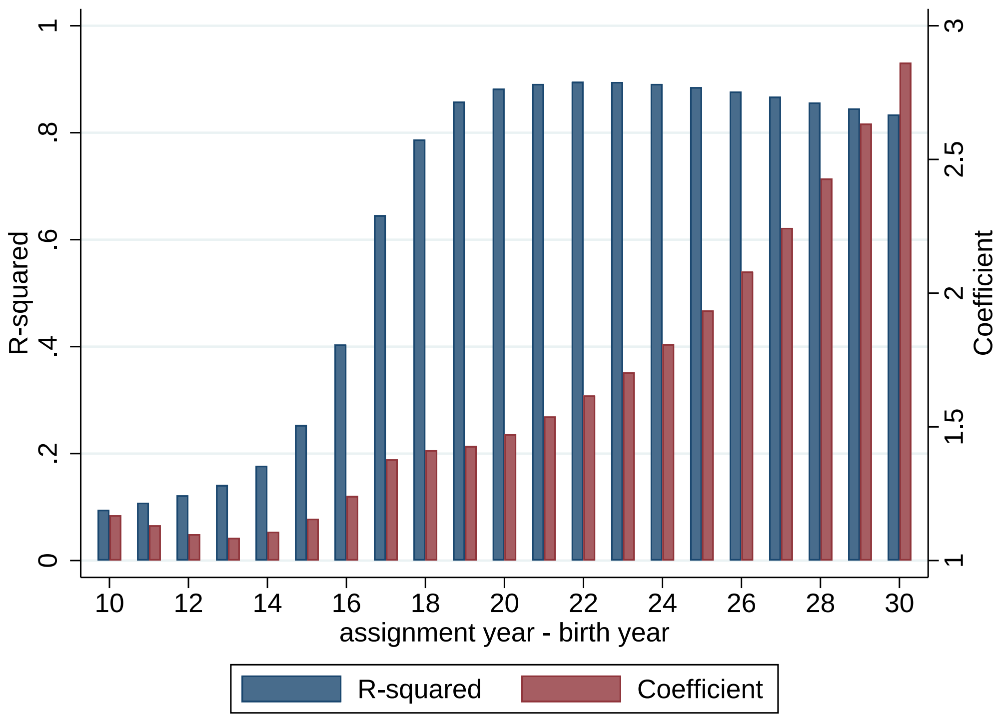
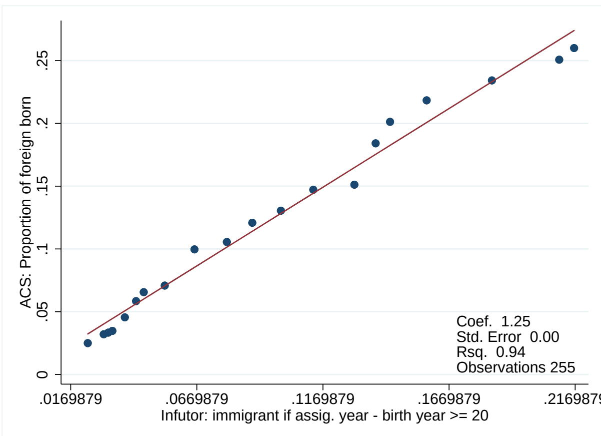


Figure A.8

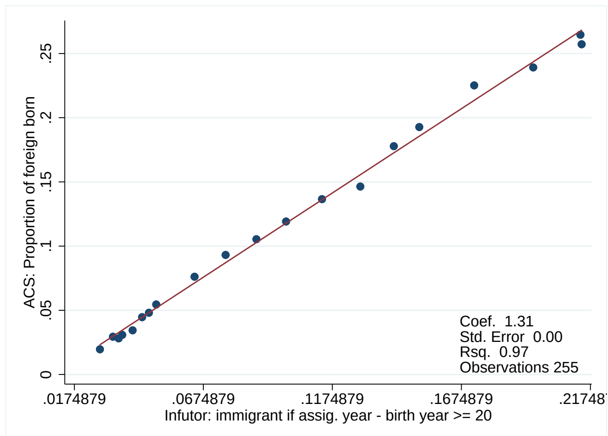
Validation with the ACS by selected age bins in 2005

Binscatters of regressing the proportion of immigrants in the State by Age level in the ACS against the same proportion in Infutor using our immigrant classification (immigrant being everyone who arrived in the U.S. after they were 20 years old) for each year and age bins. Each age bin had a sperate regression. All regressions were weighted by the number of individuals in each State and Age level. Data comes from *Infutor*, only individuals with a SSN number and a birth year.

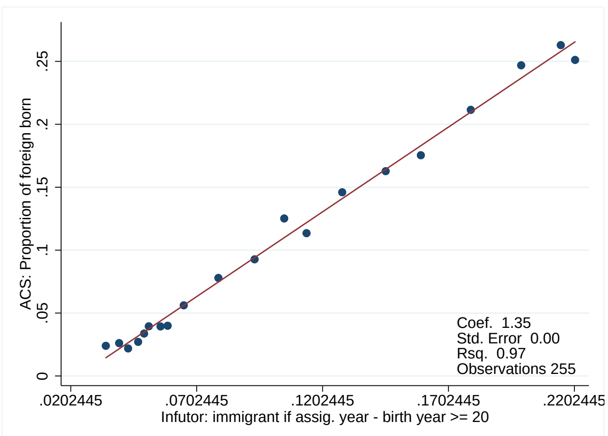
(a) **40-44 years old**



(b) **45-49 years old**



(c) **50-54 years old**



(d) **55-59 years old**

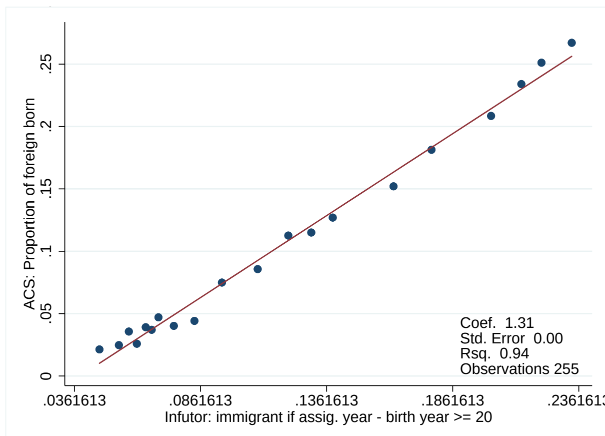
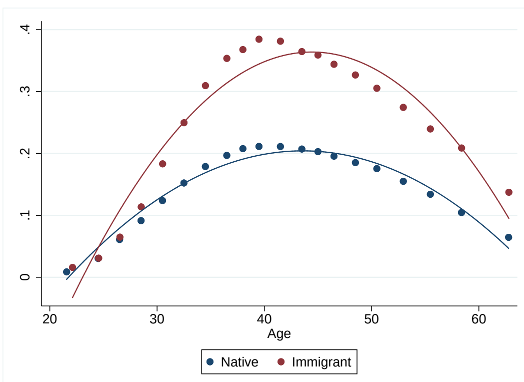


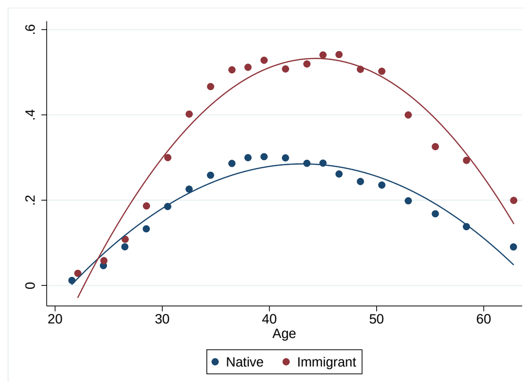
Figure A.9
Productivity over the Life Cycle - First Patent in 1990s

Categories are: (a) total number of patents per year; (b) citations,calculated over a three year horizon to avoid truncation issues, normalized by the average number of citations in a given technology class year (the year in which all patents were applied); (c) number top patents per year, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year; (d) share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only. Only individuals who applied for their first patent between 1990 and 1999.

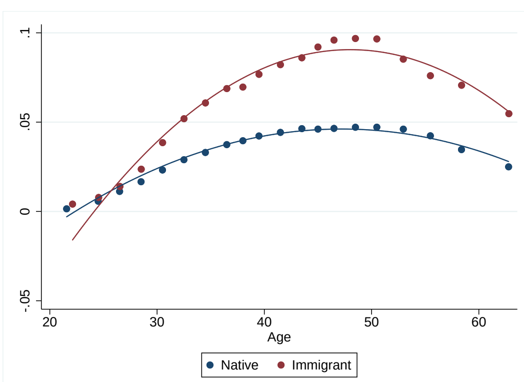
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

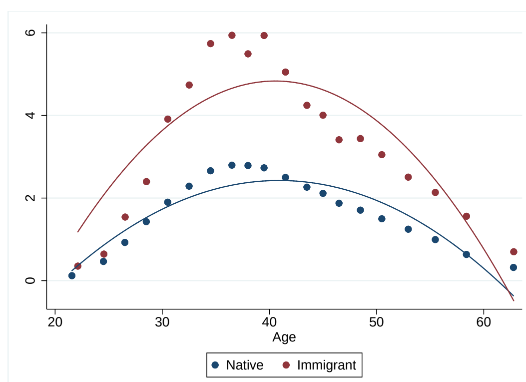
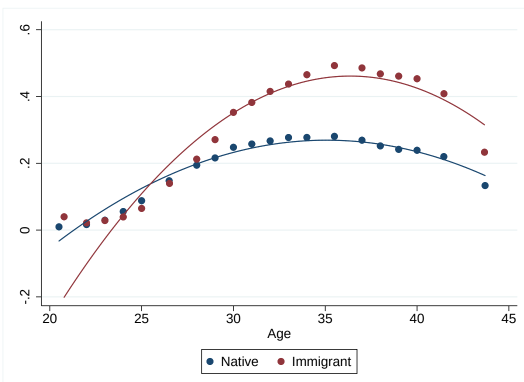


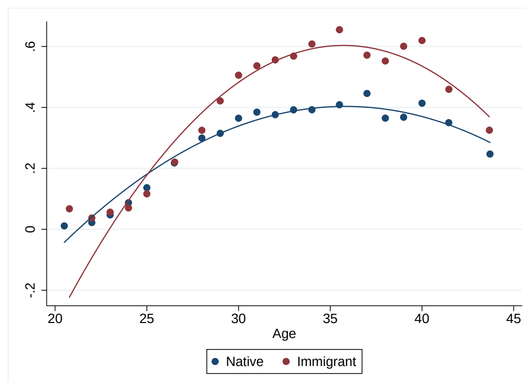
Figure A.10
Productivity over the Life Cycle - 1970s Year of Birth

Categories are: (a) share of overall number of patents; (b) citations,calculated over a three year horizon to avoid truncation issues, normalized by the average number of citations in a given technology class year (the year in which all patents were applied); (c) share of top patents, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year; (d) share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only. Only individuals born between 1970 and 1979.

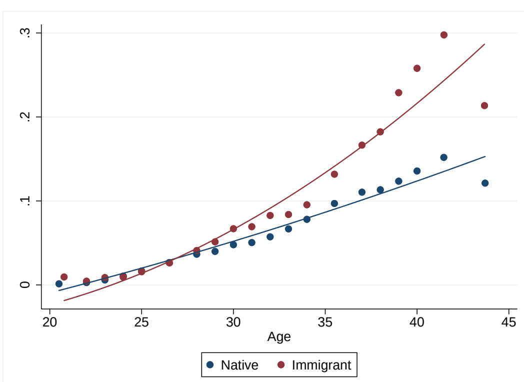
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

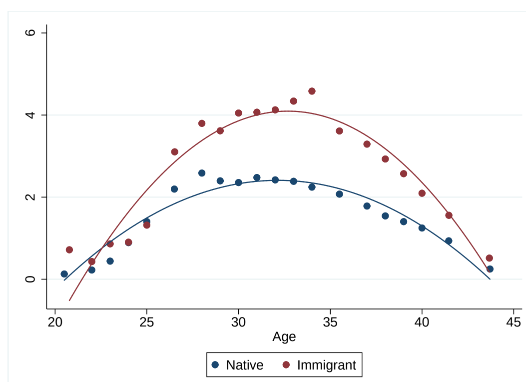
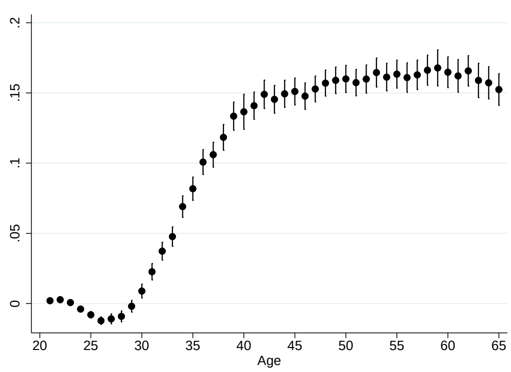


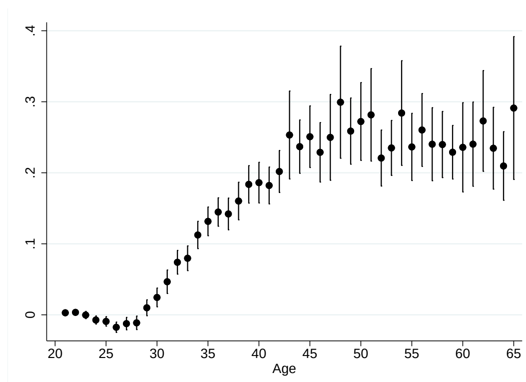
Figure A.11
Productivity over the Life Cycle - Regressions

Regression includes: individual FE, Year FE, age interacted with immigrants FE. The dependent variables are: (a) overall number of patents (b) overall number of citations first normalized by the average number of citations in a given technology class year (the year in which all patents were applied) and then added over a three year horizon to avoid truncation issues; (c) overall number of top patents, where a top patent is defined as a patent that is in the top 10% of citations in a given technology class and year. (d) Patent value calculated based on stock market reaction to patent approval using the KPSS measure. This measure is available for publicly traded firms only.

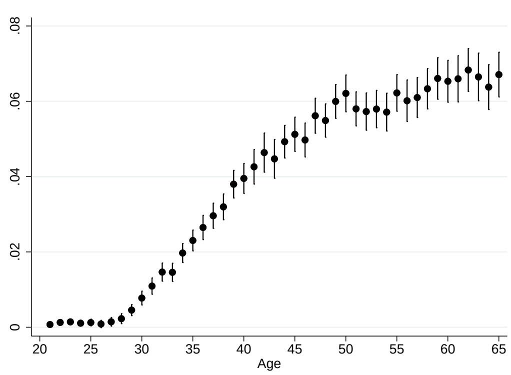
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

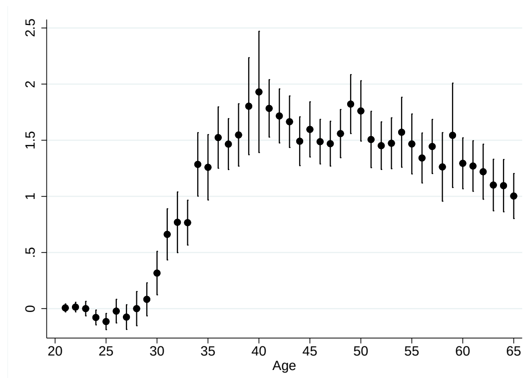


Figure A.12
Assimilation over the Life Cycle - First Patent in 1990s

Average share of immigrants in the patenting team. Only inventors who applied for their first patent between 1990 and 1999.

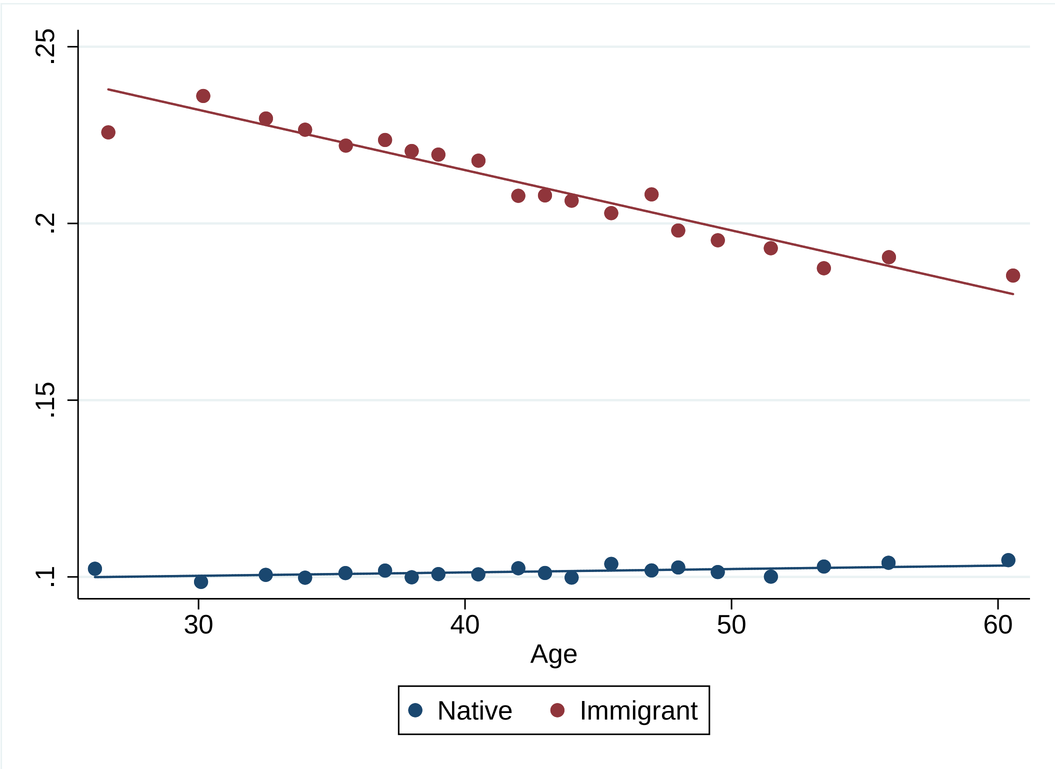
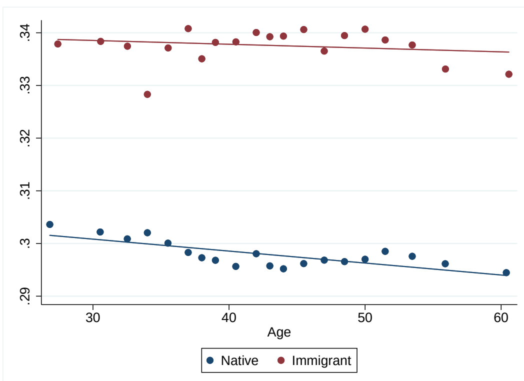


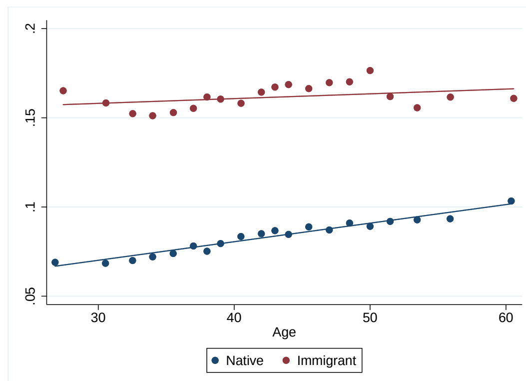
Figure A.13
Global Knowledge Diffusion - First Patent in 1990s

Citations are calculated using a three year horizion to avoid trucation issues. Categories are: (a) share of foreign patents that were cited by the inventor in their patents; (b) share of patents in which a foreign inventor is one of the co-authors in a given year; (c) share of foreign patents that cited one of the inventors patents. Only inventors who applied for their first patent between 1990 and 1999.

(a) **Share of Foreign Backward Citations**



(b) **Share of patents collaborated with foreign inventors**



(c) **Share of foreign forward citations**

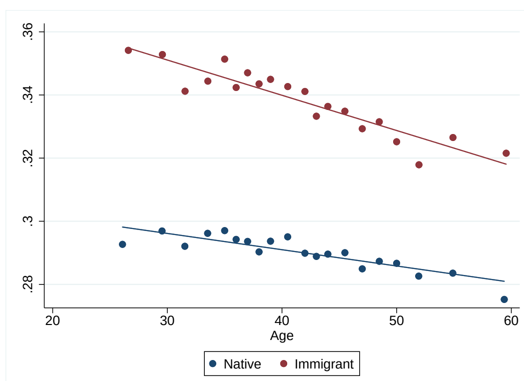


Figure A.14

Assimilation over the Life Cycle - 1970s Year of Birth

Average share of immigrants in the patenting team. Only individuals born between 1970 and 1979.

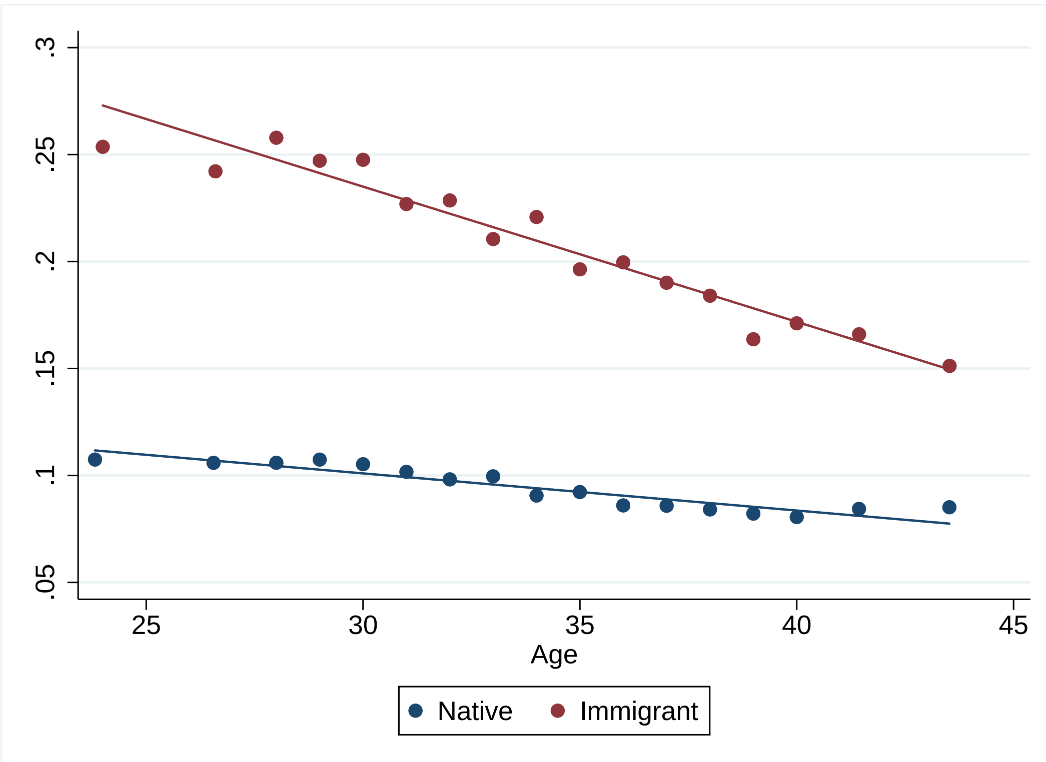
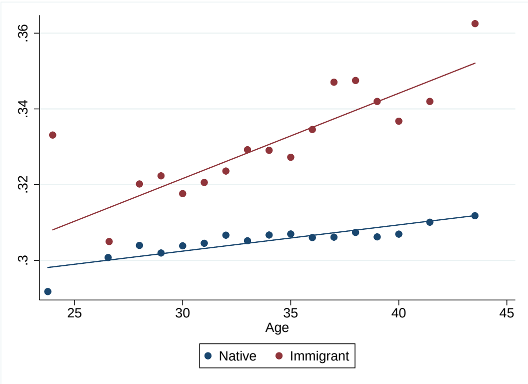


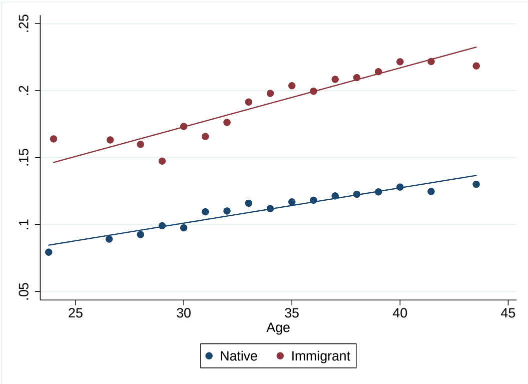
Figure A.15
Global Knowledge Diffusion - 1970s Year of Birth

Citations are calculated using a three year horizion to avoid trucation issues. Categories are: (a) share of foreign patents that were cited by the inventor in their patents; (b) share of patents in which a foreign inventor is one of the co-authors in a given year; (c) share of foreign patents that cited one of the inventors patents. Only individuals born between 1970 and 1979.

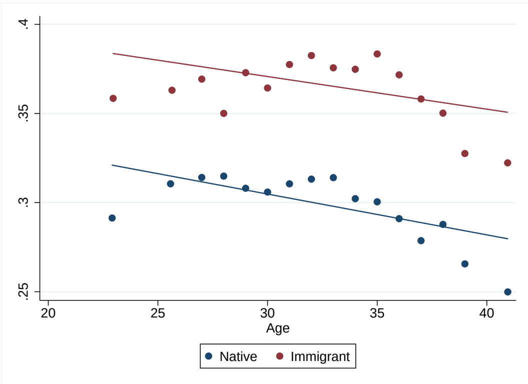
(a) **Share of Foreign Backward Citations**



(b) **Share of patents collaborated with foreign inventors**



(c) **Share of foreign forward citations**



C Tables

Table A.1
Comparing Matched and Unmatched Samples

This table shows summary statistics of the final inventor panel ranging from 1976 to 2012, for inventors that appear on Infutor against inventors that don't appear on Infutor. *Number of Patents* is defined as the number of patents applied for by an inventor during the period. *Total Citations*is the total number of citations received by an inventor. *Total adjusted citations*is citations normalized by the average number of citations in a given technology class year (the year in which all patents were applied). *Total value created* is the share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms only. *Top patents* is defined as a patent that is in the top 10% of citations in a given technology class and year.

	Mean	Median	Top90	StdDev	Obs
Number of Patents					
Matched to Infutor	4.30	2.00	9.00	9.96	944080
Not Matched to Infutor	3.84	1.00	8.00	9.96	484585
Total Adjusted Citations					
Matched to Infutor	5.72	1.18	11.82	26.75	944080
Not Matched to Infutor	5.17	1.02	10.16	26.00	484585
Top Patents					
Matched to Infutor	0.89	0.00	2.00	3.11	944080
Not Matched to Infutor	0.84	0.00	2.00	2.97	484585
Total Value Created					
Matched to Infutor	95.84	20.01	201.36	353.46	406298
Not Matched to Infutor	80.36	15.75	160.51	323.85	196221

Notes: Data comes from *Infutor*, only individuals with a SSN number and a birth year.

Table A.2
Cases without Assignment Year

This table shows the proportion of cases without a SSN assignment year in the *Infutor* sample, only individuals with a SSN number and year of birth. *U.S. territories* are area numbers used in Puerto, Rico, Guam, America Samoa and the Philippines; *not issued area* are area codes that were never issued until 2011; *not valid area* are area codes 000 and 666; *group 00* are group numbers 00; *railroad* are area codes that were used by railroad workers; *ITIN* are area numbers used for ITIN; *EaE* are area numbers used at the Enumeration at Entry program by the State Department; *not issued group* are group numbers that were never issued until 2011. All the information comes from the SSA.

	Number of obs.	Prop. of Special Cases	Prop. of total obs.
US territories	1,018,211	55.817%	0.548%
Not issued area	109,536	6.005%	0.059%
Not valid area	2,817	0.154%	0.002%
Group 00	11,809	0.647%	0.006%
Railroad	177,904	9.752%	0.096%
ITIN	95,183	5.218%	0.051%
EaE	4,675	0.256%	0.003%
Not issued group	404,061	22.150%	0.217%
Total special cases	1,824,196	100.000%	0.981%
Total observations	185,906,324		100.000%

Notes: Data comes from *Infutor*, only individuals with a SSN number and a birth year.

Table A.3
Inventor Death Infutor Sample

This table shows the diff-diff estimates of the inventor death sample, only inventors that were matched to *Infutor*. The sample and all variables are as defined in table 4. Standard errors appear in parentheses and are clustered at the deceased inventor level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Inventor Deaths

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Number of Patents	Number of Patents	Number of Patents	Top Patents	Top Patents	Top Patents
	All	Immigrants	Natives	All	Immigrants	Natives
After Death Real	-0.0461**	-0.145***	-0.0331	-0.0182***	-0.0723***	-0.0110
	(0.0198)	(0.0533)	(0.0213)	(0.00663)	(0.0203)	(0.00699)
Control Post-Mean	0.49	0.68	0.46	0.14	0.23	0.13
Percent Change	-09%	-21%	-07%	-13%	-31%	-08%
Observations	587307	69529	517778	587307	69529	517778
Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes

Panel B: Inventor Deaths

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Adjusted Citations	Adjusted Citations	Adjusted Citations	Econ Value	Econ Value	Econ Value
	All	Immigrants	Natives	All	Immigrants	Natives
After Death Real	-0.0639	-0.562***	0.00180	-2.627***	-9.539***	-1.718*
	(0.0724)	(0.213)	(0.0768)	(0.867)	(2.581)	(0.919)
Control Post-Mean	0.70	1.17	0.63	6.84	12.86	6.05
Percent Change	-09%	-48%	0%	-38%	-74%	-28%
Observations	587307	69529	517778	587307	69529	517778
Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes