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

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

March 20, 2023

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 more than 230 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 immigrants represent 16 percent of all US inventors, but produced 23 percent of total innovation output, 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 immigrant inventors create especially strong positive externalities on the innovation production of their collaborators, while natives have a much weaker impact. A simple decomposition illustrates that immigrants are responsible for 36% of aggregate innovation, two-thirds of which is due to their innovation externalities on their native-born collaborators.

*Shai Bernstein is with Harvard University and NBER, Rebecca Diamond is with Stanford University and NBER, Timothy McQuade is with UC Berkeley, Abhisit Jiranaphawiboon is with Stanford University, and Beatriz Pousada is with Stanford University. We thank seminar and conference participants for helpful comments and feedback. The authors have obtained IRB approval from Stanford University before conducting the analysis.

1 Introduction

Innovation and technological progress is considered to be a key determinant of economic growth (Romer, 1990; Aghion and Howitt, 1992; Jones, 1995). 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 as US-born college grads (Hunt and Gauthier-Loiselle, 2010).

At a broader level, we do not have an aggregate estimate of how immigrants contribute to US innovation. One of the key reasons for that is the lack of comprehensive data that enables identifying directly who are the immigrant inventors, how their productivity differs from native inventors, and thus estimate their contribution to aggregate innovation. 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 1990 through 2016. Immigrants, however, produce 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 between 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 the number of native/immigrant collaborators other inventors have access to. We find collaborating with immigrants leads to especially large productivity gains for inventors, relative to collaborating with US natives. A simple decomposition of aggregate innovation since 1990 illustrates the disproportionate contribution of immigrants, and the importance of indirect spillover effects between immigrants and natives.

Our analysis relies on the Infutor database, which provides the exact address history of more than 300 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 five digits of their Social Security numbers. Our methodology infers immigrant status by combining the first five digits of their Social Security Number (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.[2] 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

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

[2] This method has been used to identify immigrants in prior work by Doran et al. (2022) to study H1-B visa supply of firm hiring and Yonker (2017) to study immigrant CEOs.

multiple specifications.[3]

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 from 1990-2016, 16% of all US-based inventors are immigrants that came to the United States when they were 20 years old or older. 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 available for publicly traded firms and imputed for private firms, we find that the immigrants have generated 25% of the aggregate economic value, an increase of over 50% relative to their share of the inventor population.

The contribution of immigrants to US innovative output is not particularly concentrated in specific sectors. We find that immigrants generate over 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.[4] However, immigrants diverge from natives when reaching 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.

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 be selected based on their innate ability, we 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 to 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 their higher shares of backward

[3] 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.

[4] 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. (2014) for a survey), reflecting the necessary time to accumulate relevant human capital.

foreign citations. Immigrants are also about twice as likely to collaborate with foreign inventors, relative to native inventors. Finally, foreign inventors are about ten percentage points more likely to cite the patents of US-based immigrants relative to patents of 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. By contrast, we find that throughout their careers, 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. 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.[5] 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 premature death of a co-author. This form of identification strategy is becoming increasingly common in the literature (Jones and Olken, 2005; Bennedsen et al., 2020; Azoulay et al., 2011; Nguyen and Nielsen, 2010; Oettl, 2012; Becker and Hvide, 2013; Isen, 2013; Fadlon and Nielsen, 2021; Jaravel et al., 2018).

Overall, we find that premature death leads to a 10 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 17%, while a US-born inventor's death lowers productivity by approximately 9%. These gaps are large and persistent, and take place across all of our measures of innovative productivity.

To explore potential mechanisms driving these differential productivity effects of immigrants, we estimate a detailed heterogeneous treatment effects model. Even after controlling for a host of observable characteristics that could explain heterogeneous treatment effects, the productivity spillover gap between natives and immigrants persists and, in fact, grows even larger. To further understand why the gap widens, we estimate a Gelbach (2016) decomposition of the difference

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

based on ten dimensions of treatment effect heterogeneity. While the decomposition illustrates a meaningful treatment effect heterogeneity that is consistent with economic intuition, the gap persists. Our inability to reduce the gap provides strong evidence that there is indeed something special and unique about the immigrant inventor, an immigrant "secret sauce," that drives large productivity spillovers on their US-based co-authors,which cannot be easily replicated.

Finally, we use a simple framework, combined with our causal estimates of collaborator spillovers, to quantify the share of aggregate innovation which can be attributed to immigrants, both through their direct output and indirect knowledge spillovers. We conclude that 36% of total US innovative output, since 1990, can be ascribed to US immigrants, despite only making up 16 percent of the inventor workforce and only directly authoring 23% of patents. This additional 13 percentage points of innovation, over and beyond immigrants' direct output, is due to immigrants' substantial collaboration externalities on native-born inventors. Moreover, the decomposition also highlights the importance of the two-way spillovers between immigrants and natives, where over a third of US innovation can be attributed to this source, highlighting the importance of diversity, of combining inventors with different knowledge and backgrounds to push the innovation frontier.

Related Literature Our paper contributes to several bodies of literature. It is most directly linked with a growing literature that studies the role of high-skilled immigration in driving US-based innovation. One strand of this literature seeks to descriptively measure the contribution of immigrants to US innovative output. In itself, this is a challenging task due to the lack of immigrant status in the raw patenting data. Hunt and Gauthier-Loiselle (2010) uses survey data from the 2003 National Survey of College Graduates to show immigrants patent at double the rate of US natives. Furthermore, Hunt (2011) shows that the most innovative immigrants enter the US on a student or temporary work visa, as compared to alternative visa entry points. While we do not have data on visa status, we can identify immigrants directly in the patent data. This gives us panel data on the patenting behavior of each inventor, as opposed to the cross-sections available in survey datasets. We add to the findings of Hunt and Gauthier-Loiselle (2010) and Hunt (2011) by creating a new panel dataset tracking immigrants and native patenting over time, leveraging the Infutor name and residential location data to disambiguate inventor name-city pairs appearing on each patent. [6] We document the lifecycle evolution of both immigrant and native inventors. Our data also show that at most one-third of the immigrant-native innovation gap can be explained by differential sorting across geography and technology class. We find immigrants are unique in their connectivity to global knowledge networks, as measured by citation networks.

A few papers have relied on ethnic-name databases to classify scientists with names associated with specific foreign countries as likely immigrants (Kerr, 2008b,a, 2010; Foley and Kerr, 2013), Kerr (2008b) uses this approach to show that Chinese and Indian contributions to US technology

[6] This provides an alternative procedure to Balsmeier et al. (2015) for disambiguating inventors across patents. Balsmeier et al. (2015) links patents to the same inventor based on similarity of the inventor name, co-author names, text description of assignees, patent class, and inventor city and state. Thus, inventors who patent in multiple distinct technology classes or who migrate across geography are unlikely to get their patents accurately linked. Since we use Infutor as a backbone, we do not have to compare patent similarity, but rather to whether the Infutor data lists the inventor as living in the same city listed on the patent at the time of application.

formation increased dramatically during the 1990s and that ethnic innovation is concentrated in high-tech sectors. Kerr (2010) shows that ethnic inventors are more spatially concentrated than US natives and may thus contribute more to agglomeration forces. While this approach does allow one to create ethnicity flags in the patenting data, as pointed out by Kerr (2008b), the method cannot differentiate foreign-born individuals from US natives with ethnic names, nor does it allow one to track the same inventor across patents. In contrast, the panel data we construct includes a immigrant flag based on age at SSN assignment. While our method is unable to identify immigrants who move to the US at children, we accurately measure the adult immigrant population, regardless of country-of-origin.

Our paper contributes to a second strand of the literature focused on the external effects of immigration on US-born inventors and aggregate innovation. This literature has mixed results that differ depending on the context. Recent work by Doran et al. (2014) shows that firms who win H1-B lotteries are no more innovative than losers, and Borjas and Doran (2012) find that increased inflows of Soviet mathematicians to the US lowered the productivity of US-born mathematicians. In contrast, Chellaraj et al. (2008), Hunt and Gauthier-Loiselle (2010), Kerr and Lincoln (2010), and Burchardi et al. (2021) find evidence that an increase in supply of immigrant inventors does not crowd out US-born innovation, and provide some evidence of crowd-in. We focus on a particular channel through which immigrant and native inventors can have external effects on each other: collaboration externalities. We provide quasi-experimental evidence that immigrants make US natives more productive and contribute at a higher rate to innovation collaboration externalities. Indeed, nearly 40% of total innovation can be attributed to the effects of collaboration externalities (between both immigrants and natives). However, since we do not capture every channel through which immigrants may have external effects on native inventors, we cannot quantify to total effects of a change in the aggregate supply of immigrant inventors on US-born inventor productivity.

Finally, another strand of the literature provides a historical perspective (Moser et al., 2014; Akcigit et al., 2017; Sequeira et al., 2020) Moser et al. (2014) shows that Jewish chemists fleeing the German Nazi regime to the US lead to increased US-based innovation in chemistry. Akcigit et al. (2017) use 1880-1940 Censuses linked to patent records and show that immigrant inventors were more productive over the lifecycle than US-born inventors. We add to this literature by quantifying the contribution of high-skilled immigrants to overall US innovative output over the past 30 years, while quantifying the indirect collaboration spillovers between immigrants and natives, highlighting the benefits of diversity to innovation.

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 and in Section 6 we provide a simple decomposition to explore the aggregate contribution of immigrants and and the importance of 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 five 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 300 million US residents.[7] 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, 83% of the data provides the first five digits of the individual's social security number. This data was first described and made use of by Diamond et al. (n.d.).

This data appears to be highly representative of the overall US adult population.[8] 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.1 illustrates this relationship for the year 2000. Using the variation across counties, we find each additional person in Infutor predicts an additional 1.28 people living in that county, according so the 2000 Census. This implies Infutor covers 78% of the overall adult US population. 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 1990 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

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

[8]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.

each inventor on the patent lives.[9]

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 link each inventor listed on each patent to the Infutor data using name and city and state of residence at time of patent application. See Appendix A for details. We are able to merge 70% of patent-inventors to an Infutor record. Since Infutor only covers 78% of the US population, this implies a merge rate of 90% within the Infutor sample. As a comparison, Jaravel et al. (2018) merge US inventors to the IRS tax records from 1996-2012 and obtain a merge rate of 85%.[10] Using this procedure thus gives us a panel of inventors from 1990-2016, whereby in each year, we have data on any patents an inventor applied for (and was ultimately granted).

In the complete patent dataset, there are roughly 880,000 unique inventors over the 1990-2016 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 an 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 1990 to 2016 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).[11] 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

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

[10] An alternative method to linking each inventor-patent pair to Infutor would be use previously made inventor IDs produced by Balsmeier et al. (2015). These inventor IDs are created through an algorithm that combines inventor names, locations, co-authors, associated firms, and patent classifications to create an inventor identifier, using only the patent data. Since Balsmeier et al. (2015) does not have the Infutor data to rely to disambiguate inventors, their methods have a hard time linking patents from different fields to the same inventor, even if the inventor really did patent in different fields.

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

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 three-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 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). Following Kline et al. (2019), we impute the economic value for private firms using the relationship between KPSS value among publicly traded firms and patent application and assignee-level covariates. The imputation regression is shown in Appendix Table A.1.

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, which we use as our baseline cutoff. 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 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.[12] A second issue is that we will miss illegal immigrants, as they would not have SSN. However, this is likely less of an issue

[12]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.

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.[13] 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.[14] Within each area number, the next two digits, the group numbers, were assigned 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.[15]

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

[14] 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 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).

[15] Group numbers were assigned in a non-consecutive order: first odd-numbers from 01 to 09, second even numbers

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.[16]

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.[17] 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.[18]

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.[19] We also explore alternative, more conservative classifications of immigrants, requiring gaps of 21 to 25 years between the SSN assignment year

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.

[16]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 (Figure A.2).

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

[18]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.

[19]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.

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.[20] To do so, we first geocode 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.4 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.5 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.[21]

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. The ACS allows us to not only incorporate individuals age but also, importantly, identify the age in which immigrants arrived to the US. In principle, this allows us to identify in the ACS exactly those immigrants we propose to identify in Infutor. Due to confidentiality restriction, we cannot work with the data at the county level. 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

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

[21] In Figure A.6 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.

and immigrated after the age of 20, as reported by the ACS, against the same statistic constructed through Infutor.

Figure A.7 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 have about 650,000 unique inventors that have non-missing SSNs and birth dates. We first see that the productivity distribution for inventors is highly right-skewed. The median inventor has two patents, four citations, and approximately one adjusted citation over the course of a career. The median inventor also generates $27 million of economic value, as measured by stock price reaction measure taken from (Kogan et al., 2017), and no top patents. The mean inventor, by contrast, has 4.88 total patents, 24 total citations, 6.73 adjusted citations, and 1.26 top patents. Most significantly, the mean inventor is associated with patents generating $91 million of economic value. Note that for patents with multiple co-authors we apportion the patents output equally across all inventors. E.g. if a patent has 2 inventors, this would only count as half a patent of output for each inventor.

This right-skewness is also apparent at the patent level. The median patent has 1 citation, 0.42 adjusted citations, and generates $11.83 million in economic value. The mean patent has 4.5 citations, 1.29 adjusted citations, and generates $18.62 million of economic value. The table also reports that the mean age of an inventor filing a patent is 47 years (median is 46).

Finally, Table 1 provides some basic summary information on the demographics of inventors in our sample. 11 percent of the inventors in our sample are female and 17 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.[22]

Given that we find 16% of inventors in our sample are immigrants, the next natural question is what was the overall share of US innovative output between the years of 1990 to 2016 that 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 23% of all patents produced over the time period of our sample. Remarkably, this represents a 43% 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 23%, suggesting that the higher production of patents by immigrants is not coming at the cost of the lower quality. 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 25% of top patents in our sample period.

We next explore the share of total economic value that immigrants have generated over the last four decades.[23] We find that the immigrants have generated 25% of the aggregate economic value created by patents in publicly traded and private companies between the years of 1990-2016.

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 the four

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

[23] We rely on the Kogan et al. (2017) measure that captures stock market reaction to patent grants. This measure is available originally for publicly traded firms; we impute the value for private firms following Kline et al. (2019) as illustrated in Appendix Table A.1.

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 study patenting activity throughout the span of each inventor's 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. This figure plot average outcomes by age, separately for immigrants and natives. 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.[24] 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 more than 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. (2014) 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

[24] Hunt et al. (2013) also document a similar age profile of patenting for men and women, albeit with coarser data.

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, they observe 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. (2010), among others. Indeed, according to our data 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 0.093 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 0.071 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 flexible 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 0.063 more patents per year, even when accounting for these sources of differential sorting, and the effect is highly statistically significant at 1% level.

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 32% 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 0.087 more annualized citations adjusted number of patents, 0.02 more annualized top patents, and $0.95 million more in annualized economic value.

In the Appendix, we show the inverse U-shape of the innovation production function of immigrants and natives still remain when we add such controls, as well at the peak of one's career immigrants still remain significantly more productive. See Figures A.8, A.9, and A.10.

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 15%. 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, by 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 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).[25] 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.

[25] 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.

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 (relative to earlier 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 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.[26] 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.

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 28,404 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).[27] 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.

Next, we refine our sample of deceased inventors in the following ways. First, we keep only those inventors who died at the age of 60 or earlier. The goal of this restriction is to primarily capture only premature 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. In addition, we restrict our sample to deceased inventors who we can unambiguously impute their immigrant status. Applying these restrictions results in 6,043 real deceased inventors.

[26] See, for example, Jones and Olken (2005); Bennedsen et al. (2020); Azoulay et al. (2011); Nguyen and Nielsen (2010); Oettl (2012); Becker and Hvide (2013); Isen (2013); Fadlon and Nielsen (2021); Jaravel et al. (2018).

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

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 6,043 deceased inventors using this procedure. When there are multiple matches to real deceased inventors, we randomly select up to ten placebo matches to obtain a sample of one-to-many matches. Finally, we remove inventors for whom we cannot find their associated co-authors prior to death and also remove inventors who died before 1995 to ensure that we have enough pre-death periods in the difference-in-difference analysis below. We end up with 3,947 matching groups of real-deceased and 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, 51.1 years old and has filed an average of 3 patents. Ten percent of the deceased sample are immigrants. Since we match also on the ventiles of accumulated number of co-author pre-death, real and placebo deceased are balanced on that dimension as well, with 3.45 and 3.18 co-authors, respectively.

Panel (a) 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 3.97 total adjusted citations, have 0.50 top patents, have generated an average of $76 million of economic values, worked on average with a team size of 3.37 collaborators. These statistics for the placebo deceased are, respectively, 3.69 adjusted citations, 0.47 top patents, $65 million of economic value, and a team size of 3.33 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 91,964 co-inventors of the placebo deceased, whom we refer to as placebo survivor coauthors, and 15,471 co-inventors of the real deceased inventors, whom we refer to as real survivor inventors.

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 well-balanced. The surviving co-authors of real deceased are, on average, 48.3 years old. Fifteen percent are immigrants and 10 percent are female. Placebo co-authors are, on average, 46.5 years old, with 19 percent immigrants and 11 percent female. Real surviving co-authors co-patented, on average, 1.91 patents with the deceased prior to death. They have, on average, filed 8.63 cumulative patents, 1.65 top patents, and received 12.6 total adjusted citations. Placebo surviving co-authors are very similar. On average, they have co-patented 1.87 innovations with the deceased, filed 6.90 cumulative patents, 1.28 top patents, and received 9.93 total adjusted citations.[28] 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.

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.[29] This allows for arbitrary trends within the set of surviving inventors among each matched pair of real and placebo dying inventors. These additional controls give us more power. Specifically, we estimate the following OLS specification:

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

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.[30]

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

[28] 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.

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

[30] Since we match each treatment death to a set of placebo control deaths from the same year and allow for arbitrary time trends for this set of matched treat-control inventors, we do not suffer from the issues discussed in the new difference-in-differences econometrics literature (Roth et al., 2022). These issues in difference-in-differences estimation come from the use of two-way fixed effects models where the time trends depend on calendar year, not the event year of each matched treat-control sample.

death on collaborators. We also estimate this second specification by OLS:

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

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 (scaled 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 four measures of innovative productivity, we find 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.

We first focus on the annual number of patents produced. In column (1) of Table 4, we provide the estimate for all inventors, regardless of whether the deceased inventors are immigrants or natives. The coefficient β^{real} equals to -0.091 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.[31] Relative to this expected mean (reported in Table 3), the treatment effect implies that a deceased inventor lead collaborators to produce 10.6% 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 17.3% decline in patenting. By 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,

[31] 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 individual 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 individual fixed effects together into the treatment dummy, and average the calendar year FEs into the post dummy.

is only 9.4%. The dynamic treatment effects for natives and immigrants around year of death are plotted in Panels (a) and (b) of Figure 6.

In columns (4) to (6) of Table 4, we focus on the *adjusted citations* measure and find similar results. The dynamic treatment effects around year of death are plotted in Panel (b) of Figure 6. As shown in column (4), for all inventors, we find a statistically significant coefficient of -0.150, which is equivalent to a 14.3% decline in the number of adjusted number of citations 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 22.9% in adjusted citations filed by co-authors, while the effect is only 12.6% for natives. In columns (7) to (12) of Table 4 we explore two additional dimensions of innovative productivity, the number of top patents 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 Figure 7.

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 columns (1)-(3) we find that when a native co-author is lost, this results in a loss of 1.169 co-authors going forward, reflecting a decline of 1.146 native co-authors and 0.024 immigrant co-authors. As shown in columns (4)-(6), when an immigrant co-author is lost, this results in a loss of 0.787 co-authors going forward, reflecting a loss of 0.993 immigrant co-authors, partially offset by an increase of 0.206 in the number of native co-authors. These estimates will be important in our subsequent quantitative analysis decomposing the immigrant contribution to US innovation.

5.4 Mechanisms driving spillover differences

Having established differential productivity impacts of native versus immigrant inventor deaths on co-authors, a natural question is what drives this difference. For example, we have shown that immigrant inventors are more productive in their patenting output than natives. To the extent that more productive inventors have larger spillover effects, this could drive our results. Alternatively, and perhaps more interesting from a policy perspective, it may be that there is something special and unique about the immigrant inventor that drives the large productivity spillovers.

We investigate these issues by estimating heterogeneous treatment effects of inventor deaths along a number of observable dimensions of inventor characteristics in addition to estimating the effect of immigrant versus native deaths. For example, the average dying immigrant inventor is 0.48 years older than the average dying native inventor. If we allow for treatment effect heterogeneity of an inventor death based on their age, then our estimated spillover gap between native and immigrant inventors would only capture the differential effects above and beyond their age differences. To implement this, we allow for treatment effect heterogeneity based on the dying inventor's age, the dying inventor's year of death, the dying inventor's cumulative patents and citations prior to death, the dying inventor's average coauthors' ages, the dying inventor's average coauthors' cumulative patents and citations (measured in the year prior to death), the collaboration recency between the dying inventor and surviving coauthor, the number of unique prior coauthors prior to death, the number of co-patents between dying and surviving coauthors prior to death, the similarity in dying and surviving inventor's prior work as measured by the share of their own patent's backward citations to over-lapping technology classes, and the number of patents per-capita in their commuting zone of residence, as measured in the year prior to death.

In Table 6, Panel A reports the baseline treatment effect differential between natives and immigrants before controlling for observables. Panel B reports the treatment effect differential after controlling for observable treatment effect heterogeneity and Panel C reports the difference between the two. We see that after controlling for a host of heterogeneous treatment effects along many observables, the productivity spillover gap between natives and immigrants persists and, in fact, grows even larger. The productivity spillover gap increases by 0.128 for patents, by 0.121 for citations, by 0.024 for top patents, and by 1.585 for economic value.

To investigate why the immigrant-native spillover gap gets even larger we estimate a Gelbach (2016) decomposition of the difference based on the ten dimensions of treatment effect heterogeneity. Positive (negative) percentages indicate that controlling for the observable widens (shrinks) the difference. We first see from Row 3 of Panel D of Table 6 that controlling for the productivity of the dying inventor does narrow the treatment effect gap between immigrants and natives. That is, consistent with the discussion above, the treatment effect on co-authors increases in magnitude with the dying inventor's productivity. This, together with the fact that dying immigrant inventors are more productive than dying native inventors, as shown in column (5), reduces the gap in productivity spillovers. In particular, controlling for productivity at the time of death reduces the treatment effect gap by 34.2% for the number of patents, by 60% for scaled citations, by 68.8% for top patents, and by 8.9% for economic value.

We also see from Row 7 that controlling for the size of the collaboration network, measured as the number of unique co-authors, reduces the treatment effect gap between immigrants and natives across all productivity measures. Dying inventors with larger collaboration networks have larger productivity spillover effects on their co-authors and, as shown in column (5), dying immigrants inventors have larger networks than dying native inventors. It reduces the gap by 15% for the number of patents, by 16.9% for scaled citations, by 9.5% for top patents, and by 17.3% for economic value. The Gelbach (2016) decomposition shows that controlling for knowledge gap and the commuting zone number of patents reduces the treatment effect gap between immigrants and natives when productivity is measured according to economic value.

As we see in Panel C, however, the treatment effect gap actually grows *larger* when we match on this full set of observables. This is largely driven by controlling for average co-authors' productivity at the time of death. Intuitively, the productivity spillover effects are weaker on more productive co-authors. Since immigrants work with more productive co-authors, as shown in column (5), controlling for this heterogeneity widens the gap between immigrant and native treatment effects. Indeed, we see from Panel D of Table 6 that controlling for co-author productivity has large effects, amplifying the treatment effect gap substantially across all four productivity measures.

Taken together, Table 6 shows that there is meaningful treatment effect heterogeneity that is consistent with economic intuition. Moreover, immigrants and native inventors differ on key observable characteristics, such as own and co-author productivity, that generate significant treatment effect heterogeneity. We find, however, that when controlling for a battery of observable characteristics, the productivity spillover gap between immigrants and natives persists and, in fact, is larger than in the baseline results. This provides strong evidence that there is indeed something special and unique about the immigrant inventor, an immigrant "secret sauce," that drives large productivity spillovers on their US-based co-authors.

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 that 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:

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

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's 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.[32] Taking the first order Taylor expansion around a base value and rearranging gives:

[32]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 the parameters of this production function.

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

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 represents the percent decline in output due to the change in the number of (living) collaborators. These are the exact number estimated in our difference analysis, presented in Table 4. For example, we found that the death of an immigrant co-author lowered surviving collaborators' productivity by 17.3% in terms of number of patents. The right-hand side of these equations shows 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 the total number of (living) prior immigrant and native collaborators. This is estimated in Table 5. For example, the death of a native co-author leads to 13.9% decline in the number of native co-authors and a 1.5% decline in the number of immigrant co-authors.

We next take these estimates of co-author losses from Table 5 and our estimated productivity losses from Table 4 and plug them into the first-order approximation equations above to recover the parameters of the innovation production function. These estimates are in Table 7.[33] Consistent with our reduced form findings above, we find immigrant collaborators have much larger spillover effects onto their collaborators' productivity. In particular, we find $\beta_{imm} = 1.00$ and $\beta_{nat} = 0.68$.

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 0.093 more patents 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 are still more productive that natives, statistically significant at the 1% level, but would only produce 0.015 more patents per year than native inventors. Since immigrants have large co-author networks and more immigrant collaborators, they are able to be even more productive than their "raw ability" advantage over natives would indicate.

[33]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 4 and 5. First, we use the estimates of the effects of a native death on number of patents from Table 4 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 5 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.

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 an accounting decomposition of the observed innovation we see in the data. These do not represent counterfactual 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 77 percent of the total patents 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 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. Column 2 of Table 8 shows that natives' innovation would now fall to 51% of total innovation. Thus, 26 percent (77%-51%) of total US innovation can be attributed to immigrants' collaboration spillovers on their native collaborators. This further implies that 33.8% ((77-51)/77) of natives' total innovation can be indirectly attributed to their immigrant co-authors.

Second, we explore the importance of natives' indirect contribution to immigrants' production. As shown in Table 8, immigrants produce 23 percent of total innovation in the data, as 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 8, immigrants would have produced 10% of aggregate innovation, implying natives' spillovers onto immigrants' productivity accounts for 13% of aggregate innovation.

All in all, these calculations suggest that immigrants contribute directly to 10% of innovation, and their indirect contributions, through the enhancement of natives' productivity, explain 26% of innovation. Together, immigrants account for 36% 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 39 percent of total US innovation, highlighting the importance of enhancing team-specific capital through immigrant and native diversity in teams.

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. 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 the number of patents, patent citations, and the economic value of these patents. Immigrant inventors also appear to facilitate the importation of foreign knowledge into the United States, with immigrant 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 high-skilled immigrants who provide unique knowledge that contributes not only directly to innovation but also improves significantly the innovative productivity of natives, pushing forward the innovation frontier.

Figure 1: **Share of Immigrant Contribution**

Notes. Categories are: (a) share in the overall population from 1990-2016 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)-(j) share of top patents, where a top patent is defined as a patent that is in the top 50%, 25%, 10%, 5%, and 1% of citations in a given technology class and year, respectively; (k) share of patent value, calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms and imputed for private firms.

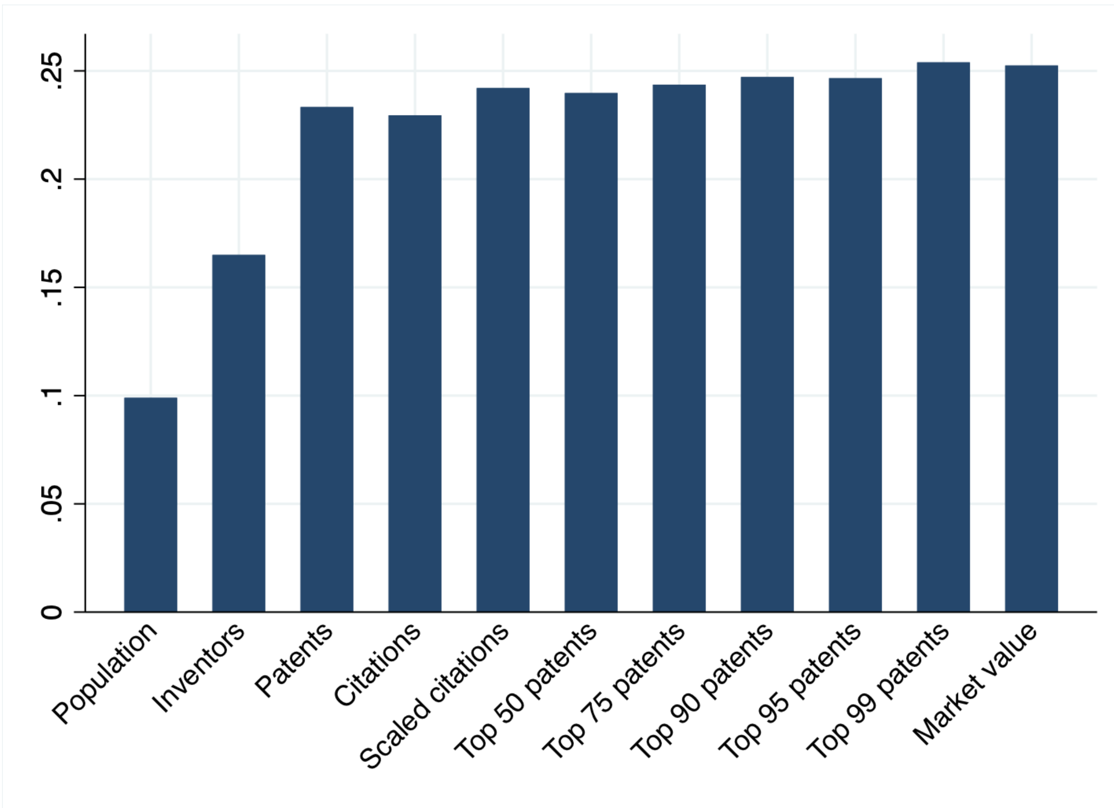
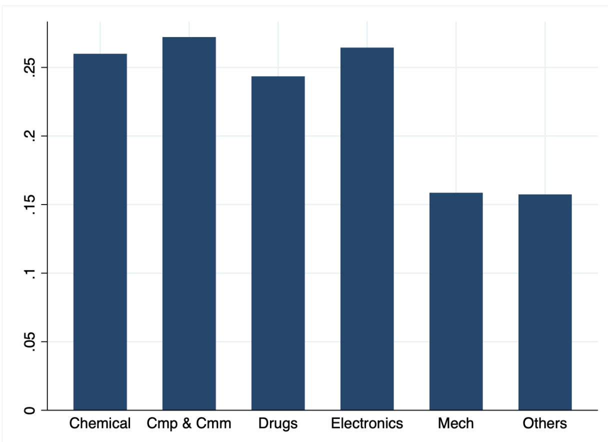


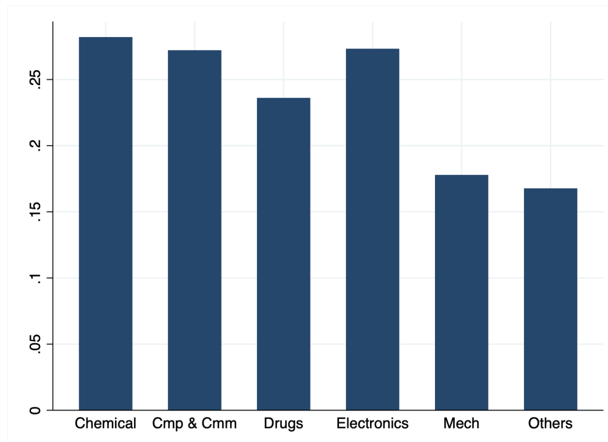
Figure 2: **Share of Immigrant Contribution across Tech Classes**

Notes. 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 and imputed for private firms.

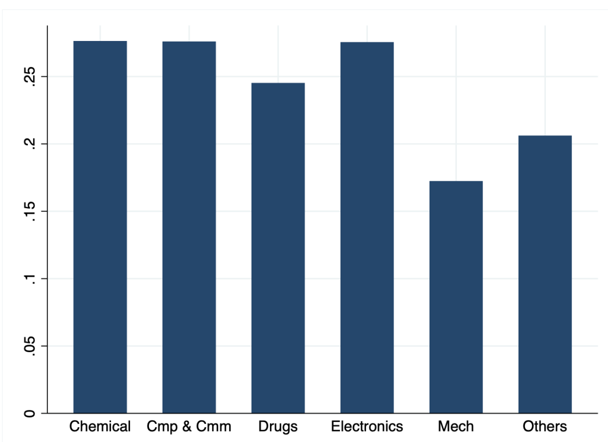
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

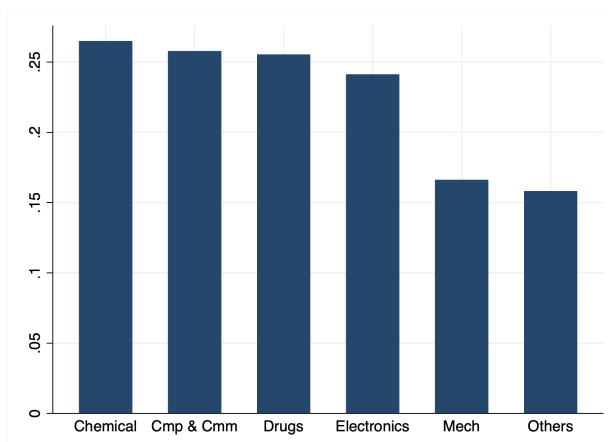
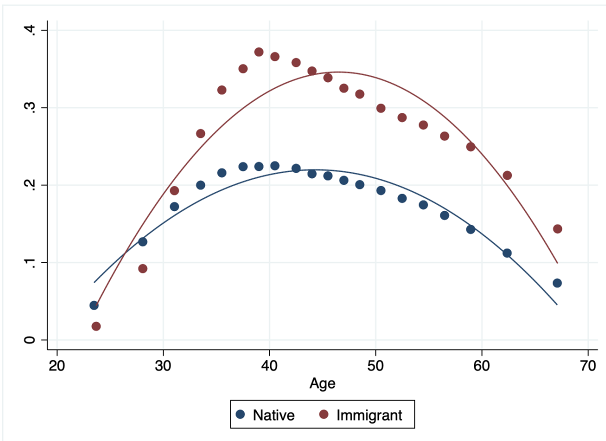


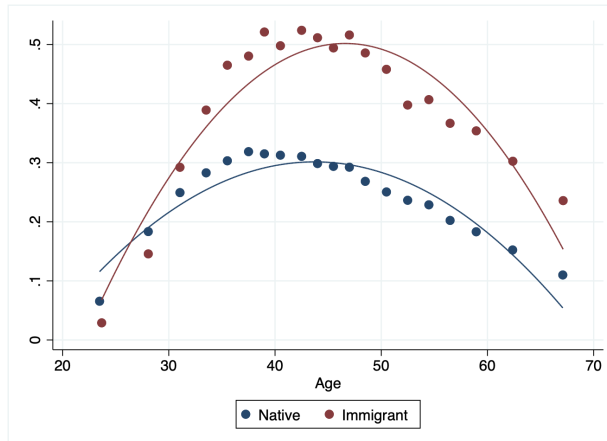
Figure 3: **Productivity over the Life Cycle**

Notes. 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 and imputed for private firms.

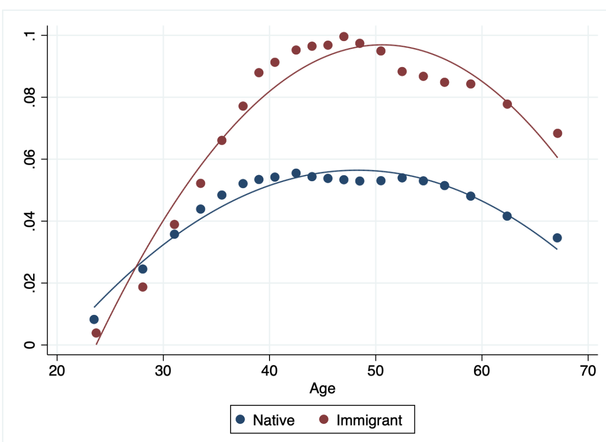
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

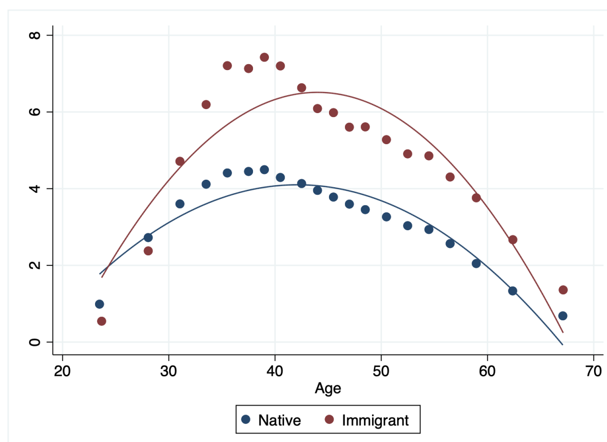
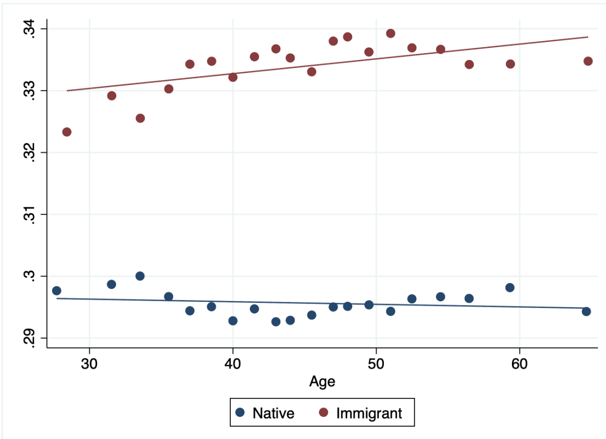


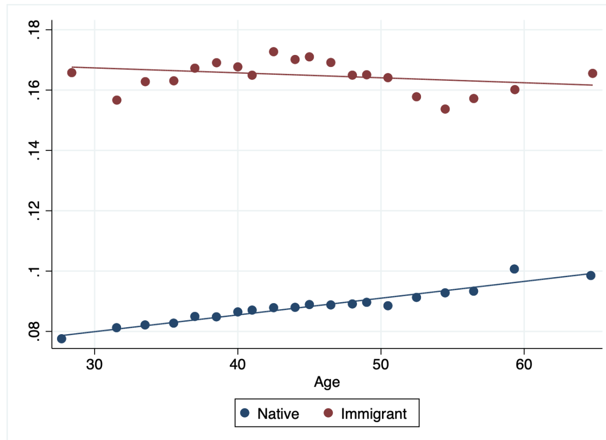
Figure 4: **Global Knowledge Diffusion**

Notes. Citations are calculated using a three year horizon to avoid truncation 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**

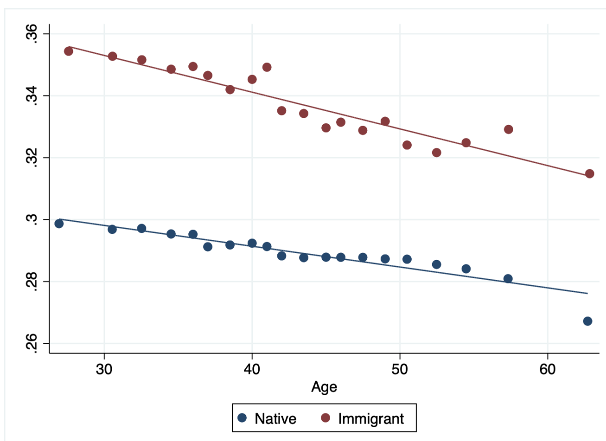
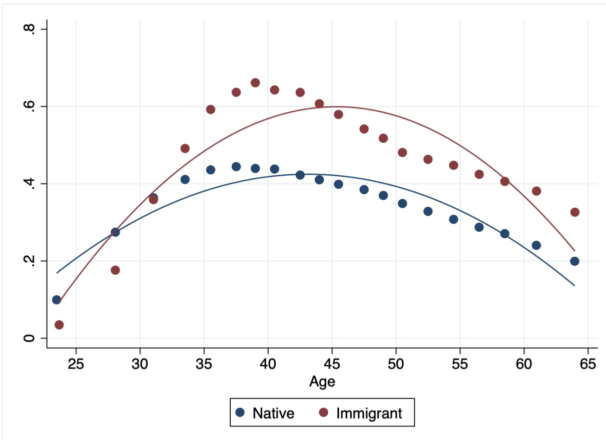


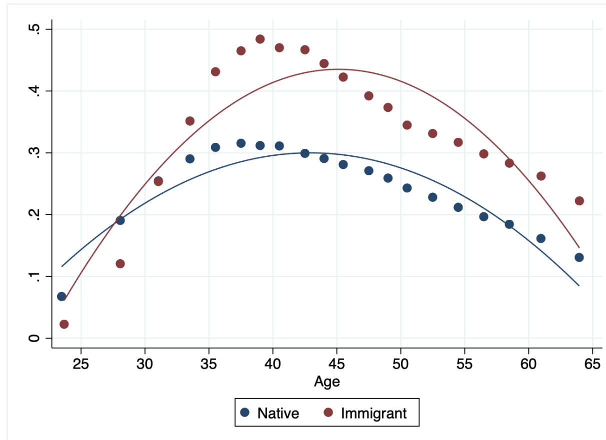
Figure 5: **Assimilation over the Life Cycle**

Notes. 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.

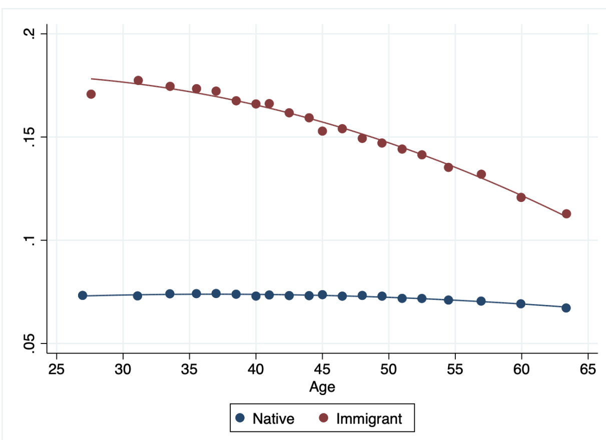
(a) **Number of unique co-authors**



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



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



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

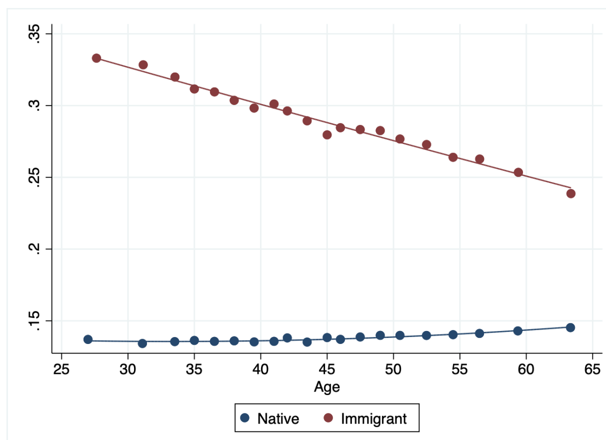
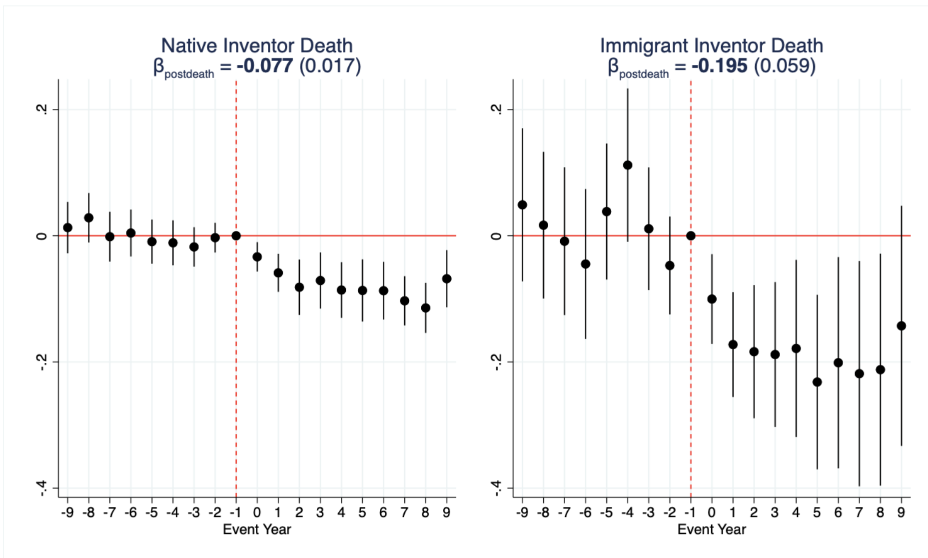


Figure 6: **Comparing immigrant and native inventors death**

Notes. 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**

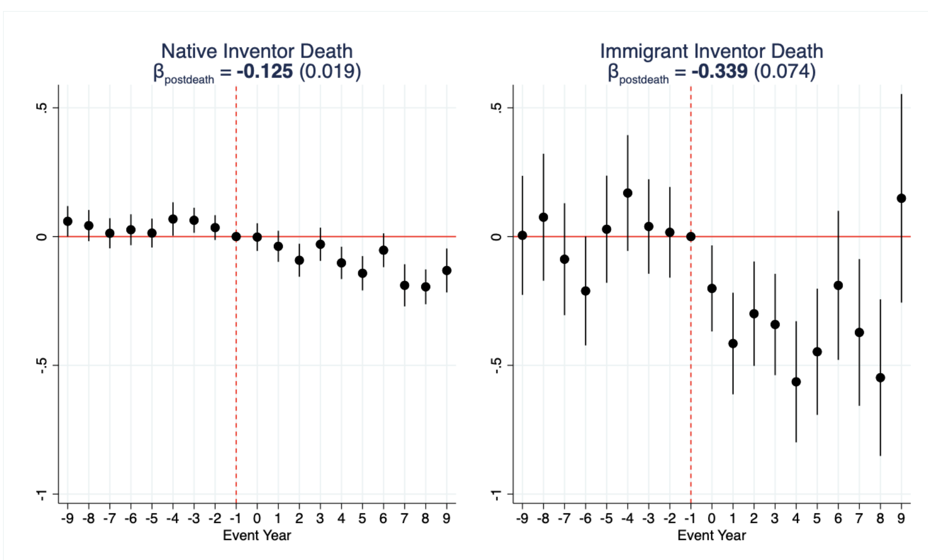
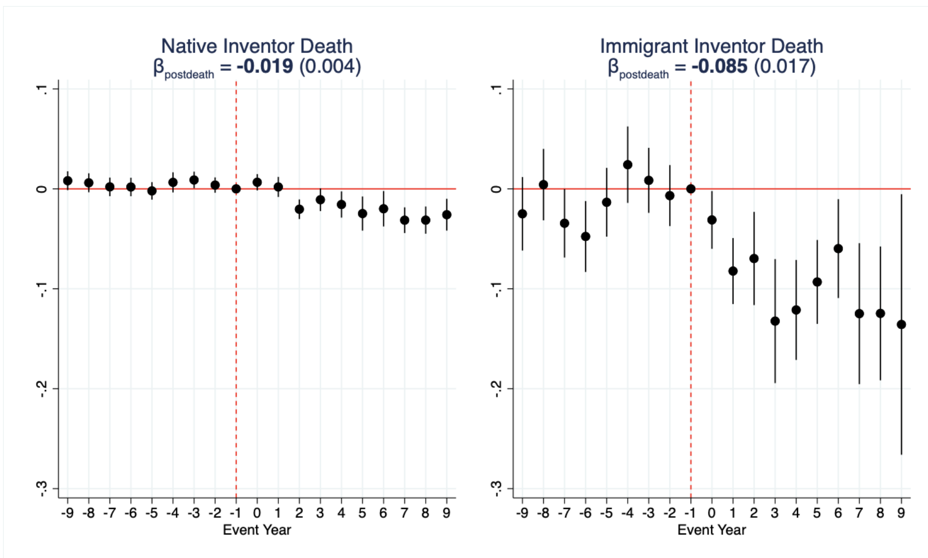


Figure 7: **Comparing immigrant and native inventors death**

Notes. 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 and imputed for private firms.

(a) **Top Patents**



(b) **Economic Value**

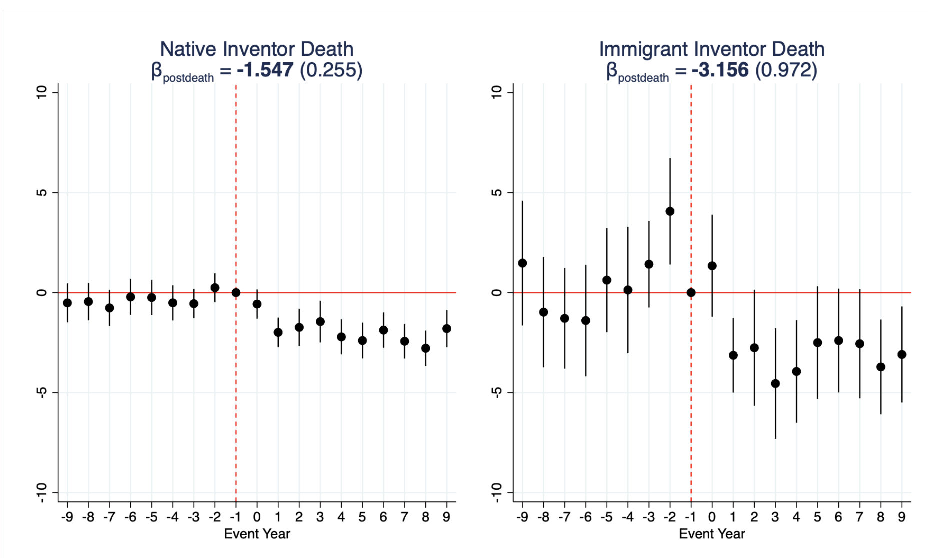


Table 1: **Summary Statistics**

Notes. This table shows summary statistics of the final inventor panel ranging from 1990 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 and imputed for private firms. *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.

	Mean	Median	Top 90%	Std Dev	# Obs
Patenting Outcomes - Inventor-Level					
Number of patents	4.88	2.00	11.00	11.67	652,832
Total citations	24.01	4.00	50.00	104.55	652,832
Total adjusted citations	6.73	1.19	13.21	35.70	652,832
Total value created	91.38	26.75	187.95	305.25	652,832
Top patents	1.26	0.00	3.00	4.40	652,832
Patenting Outcomes - Patent-Level					
Citations	4.50	1.00	11.00	10.47	1,790,161
Adjusted citations	1.29	0.42	2.84	6.34	1,790,161
Market value	18.62	11.83	35.25	37.44	1,790,161
Top patents	0.25	0.00	1.00	0.43	1,790,161
Age at application	47.09	46.00	61.00	10.66	1,790,161
Demographics of Inventors					
Female	0.11	0.00	1.00	0.31	652,832
Immigrant	0.17	0.00	1.00	0.37	652,832

Table 2: **Differential Sorting between Immigrant and Native Inventors**

Notes. This table estimates the effect of being an immigrant on inventors' productivity with different combinations of fixed effects. Precisely, we run $Y_{it} = \beta_0 + \beta_1 Immigrant_i + X\gamma + \varepsilon_{it}$, where Y_{it} denotes our outcome of interest for inventor i in year t, $Immigrant_i$ equals 1 if the inventor is an immigrant based on our measure, and X is a vector of fixed effects that we successively add to the regression. Standard errors appear in parentheses 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 and imputed for private firms. 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	0.093***	0.092***	0.071***	0.062***	0.071***	0.063***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	15,714,917	15,714,917	15,714,906	15,192,932	15,709,593	15,187,669
Year FE	yes	yes	yes	yes	yes	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
County × Tech FE	no	no	no	yes	no	no
County × Year FE	no	no	no	no	yes	yes
Tech × Year FE	no	no	no	no	no	yes

Panel B: Annual Adjusted Citations

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	0.149***	0.147***	0.100***	0.086***	0.100***	0.087***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	15,714,917	15,714,917	15,714,906	15,192,932	15,709,593	15,187,669
Year FE	yes	yes	yes	yes	yes	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
County × Tech FE	no	no	no	yes	no	no
County × Year FE	no	no	no	no	yes	yes
Tech × Year FE	no	no	no	no	no	yes

Table 2: (Continued)

Panel C: Annual Number of Top Patents

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	0.029***	. 0.030***	0.022***	0.020***	0.021***	0.020***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	15,714,917	15,714,917	15,714,906	15,192,932	15,709,593	15,187,669
Year FE	yes	yes	yes	yes	yes	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
County × Tech FE	no	no	no	yes	no	no
County × Year FE	no	no	no	no	yes	yes
Tech × Year FE	no	no	no	no	no	yes

Panel D: Annual Aggregate Economic Value

	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	1.790***	1.733***	1.213***	0.860***	1.229***	0.945***
	(0.054)	(0.054)	(0.055)	(0.055)	(0.055)	(0.056)
Observations	15,714,917	15,714,917	15,714,906	15,192,932	15,709,593	15,187,669
Year FE	yes	yes	yes	yes	yes	no
YOB FE	no	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no
County × Tech FE	no	no	no	yes	no	no
County × Year FE	no	no	no	no	yes	yes
Tech × Year FE	no	no	no	no	no	yes

Table 3: **Inventor Death Controls**

Notes. This table shows summary statistics for control variables and pre-treatment dependent variables for the real and placebo deceased and survivor inventors at the actual/counterfactual deceased year. 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. 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		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Age	51.13	53	7.05	51.13	53	7.05
Year	2,004.42	2,005	4.89	2,004.42	2,005	4.89
Immigrant status	0.10	0		0.10	0	
Cumulative patents	3	2	2.65	3	2	2.65
Co-authors	3.45	2	4.35	3.18	2	3.51
Team size	3.37	3	2.20	3.33	3	2.78
Adjusted citations	3.97	1.19	9.38	3.69	1.05	9.98
Top patents	0.50	0	1.40	0.47	0	1.48
Economic Value	76.11	23.11	265.43	64.75	20.84	197.31
Female	0.07	0		0.10	0	
Sample size	3,947			37,768		

Table 3: (Continued)

Panel B: Real vs. Placebo Co-Inventor Characteristics

	Real Deceased			Placebo Deceased		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Age	48.37	49	16.61	46.51	47	20.55
Immigrant status	0.15	0		0.19	0	
Co-patents pre-treat	1.91	1	2.26	1.87	1	2.32
Cumulative patents	8.63	3	20.46	6.90	3	15.71
Adjusted citations	12.58	3.24	35.70	9.93	2.38	27.87
Top patents	1.65	0	4.79	1.28	0	3.64
Economic Value	209.40	49.97	585.34	163.80	39.86	496.61
Female	0.10	0		0.11	0	
Sample size	15,471			91,964		

Panel C: Comparing Technologies

	Deceased Inventors		Placebo Inventors		Deceased Co-inventor		Placebo Co-inventor	
	# Patents	Share	# Patents	Share	# Patents	Share	# Patents	Share
Chemicals	2,182	0.09	16,482	0.08	21,112	0.09	77,807	0.07
Computers	2,843	0.12	24,275	0.11	25,009	0.10	118,349	0.11
Drugs	1,810	0.07	16,970	0.08	16,205	0.07	65,609	0.06
Economic Value	1,957	0.08	18,404	0.08	21,734	0.09	87,563	0.08
Female	1,929	0.08	15,486	0.07	12,953	0.05	55,208	0.05
Sample size	1,936	0.08	16,095	0.07	10,551	0.04	46,989	0.04

Table 4: **Inventor Death**

Notes. This table shows the difference-in-difference OLS 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 1. 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:	Number of Patents			Adjusted Citations		
	All	Immigrant	Native	All	Immigrant	Native
	(1)	(2)	(3)	(4)	(5)	(6)
Post Death	-0.091***	-0.195***	-0.077***	-0.150***	-0.339***	-0.125***
	(0.016)	(0.059)	(0.017)	(0.019)	(0.074)	(0.019)
Control Post Mean	0.856	1.125	0.820	1.051	1.483	0.992
Percent Change	-10.6%	-17.3%	-9.4%	-14.3%	-22.9%	-12.6%
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
R^2	0.564	0.598	0.558	0.363	0.279	0.390
Number of Deceased Immigrants	37,398	3,326	34,072	37,398	3,326	34,072
Observations	1,889,544	200,727	1,688,817	1,889,544	200,727	1,688,817
Dependent Variable:	Top Patents			Economic Value		
	All	Immigrant	Native	All	Immigrant	Native
	(7)	(8)	(9)	(10)	(11)	(12)
Post Death	-0.027***	-0.085***	-0.019***	-1.735***	-3.156***	-1.547***
	(0.004)	(0.017)	(0.004)	(0.252)	(0.972)	(0.255)
Control Post Mean	0.202	0.293	0.190	13.293	19.206	12.499
Percent Change	-13.4%	-29.0%	-10.2%	-13.1%	-16.4%	-12.4%
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
R^2	0.386	0.327	0.405	0.453	0.455	0.452
Number of Deceased Immigrants	37,398	3,326	34,072	37,398	3,326	34,072
Observations	1,889,544	200,727	1,688,817	1,882,309	197,405	1,684,904

Table 5: **Inventor Death: Impact on Number of Unique Collaborators**

Notes. This table shows the difference-in-difference OLS 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 1. 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.

Dying Inventor:	Native			Immigrant		
Number of Prior Collaborators:	Any	Native	Immigrant	Any	Native	Immigrant
	(1)	(2)	(3)	(4)	(5)	(6)
Post Death	-1.169***	-1.146***	-0.024	-0.787***	0.206	-0.993***
	(0.058)	(0.047)	(0.017)	(0.211)	(0.150)	(0.085)
Control Post Mean	9.847	8.234	1.613	10.587	6.713	3.874
Percent Change	-11.9%	-13.9%	-1.5%	-7.4%	3.1%	-25.6%
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
R^2	0.896	0.896	0.886	0.878	0.877	0.872
Number of Deceased Inventors	29,547	29,547	29,547	2,979	2,979	2,979
Observations	1,688,817	1,688,817	1,688,817	200,727	200,727	200,727

Table 6: **Decomposition of Inventor Death Treatment Effect Differentials**

Notes. This table shows Gelbach decomposition of the treatment effect differentials between immigrant and native inventor deaths for number of patents, scaled citations, top patents, and economic value in columns (1), (2), (3), and (4), respectively. Panel A reports our baseline differentials. Panel B reports the differentials after controlling for thirteen additional observables. Panel C reports the absolute difference between the estimates between the two panels. Panel D breaks this difference into percentages explained by each of the thirteen observables grouped in ten categories/rows listed below. Positive (negative) percentages mean controlling for the observable widens (shrinks) the difference. Column (5) of Panel D reports the difference in mean observables between deceased immigrants and deceased natives in the sample; to help interpret this difference, the overall means are also reported in the brackets. In Panel D, rows 1-3 consider observables featuring deceased inventors' characteristics; rows 4-5 consider observables featuring surviving coauthors' characteristics; rows 6-9 consider observables featuring interactions between deceased inventors and their collaborators pre-death; and row 10 considers the number of patents per capita in each commuting zone. We group two observables in one category for rows 3,5, and 9, but report the mean separately for each observable in these columns. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent variable:	Number of Patents (1)	Scaled Citations (2)	Top Patents (3)	Economic Value (4)	Difference in Mean Observables (5)
Panel A. Differentials *Before* Controlling for Observables					
	-0.092	-0.251**	-0.054***	-1.167	
	(0.078)	(0.123)	(0.020)	(1.749)	
Panel B. Differentials *After* Controlling for Observables					
	-0.221**	-0.372***	-0.078***	-2.752	
	(0.091)	(0.140)	(0.023)	(1.933)	
Panel C. Absolute Difference between Estimates in Panels A and B					
	0.128**	0.121*	0.024*	1.585**	
	. (0.054)	(0.065)	(0.013)	(0.635)	
Panel D. Gelbach Decomposition					
1. Deceased inventor's age	0.8%	3.2%	4.3%	2.2%	0.48 [51.12]
2. Deceased inventor's year	2.4%	-3.6%	-1.5%	1.0%	−0.24 [2004.42]
3. Deceased inventor's cumulative patents and citations	-34.2%	-60.0%	-68.8%	-8.9%	0.84 [3.00] 10.65 [22.04]
4. Average surviving coauthors' age	8.2%	7.7%	9.0%	4.1%	−0.13 [51.49]
5. Average surviving coauthors' cumulative patents and citations	115.6%	159.8%	157.7%	136.4%	2.75 [8.49] 46.55 [61.21]
6. Collaboration recency: time to most recent app pre-death	2.8%	8.0%	6.2%	4.3%	0.09 [2.76]
7. Collaboration network: number of unique coauthors pre-death	-15.0%	-16.9%	-9.5%	-17.3%	0.84 [3.92]
8. Collaboration strength: number of co-patents pre-death	14.2%	0.4%	4.1%	21.6%	3.22 [10.68]
9. Knowledge gap: backward citations and overlapping technology classes	5.1%	-3.9%	2.5%	-18.2%	−0.07 [0.53] −0.03 [0.51]
10. Commuting zone number of patents per capita	0.3%	5.4%	-4.1%	-25.2%	241.91 [707.07]

Table 7: **Innovation Production Function Estimates**

Notes. 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 impacts on number of patents as the metric of interest.

Panel A: Production Function Parameter Estimates

	(1)	(2)
	β_{imm}	β_{nat}
	1.00	0.68
Inventor Type	Immigrant	Native

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

	(1)	(2)	(3)
	Immigrants	Natives	Difference
Avg A_{it}	0.088***	0.073***	0.015***
	(0.0006)	(0.0007)	(0.002)

Table 8: **Decomposing Aggregate Innovation Output**

Notes. 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 between columns 1 and 2, representing the additional output natives produce by working with immigrants.

	(1)	(2)	(3)
Native Output	0.77	0.51	0.64
Immigrant Output	0.23	0.10	0.36
Total Output	1.00	0.61	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

References

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–51.

— **, 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. 609–26.

Audretsch, David and Maryann Feldman, "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 1996, *86* (3), 630–40.

Azoulay, Peter, Joshua Zivin, and Gustavo Manso, "Incentives and Creativity: Evidence from the Academic Life Sciences," *RAND Journal of Economics*, 2011, *42* (3), 527–54.

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.

Becker, Gary, *Human Capital Theory*, Columbia, New York, 1964.

Becker, Howard Saul, *Art Worlds*, University of California Press, 1982.

Becker, Sascha and Hans Hvide, "Do Entrepreneurs Matter?," *CESifo Working Paper No. 4088*, 2013.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, *75* (4, Part 1), 352–65.

Bennedsen, Morten, Francisco Pérez-González, and Daniel Wolfenzon, "Do CEOs Matter?," *Journal of Finance*, 2020, *75* (4), 1877–911.

Bernstein, Shai, "Does Going Public Affect Innovation?," *Journal of Finance*, 2015, *70* (4), 1365–403.

Borjas, George, *Immigration Economics*, Harvard University Press, 2014.

— **and Kirk Doran**, "The Collapse of the Soviet Union and the Productivity of American Mathematicians," *Quarterly Journal of Economics*, 2012, *127* (3), 1143–203.

Burchardi, Konrad, Thomas Chaney, Tarek Hassan, Lisa Tarquinio, and Stephen Terry, "Immigration, Innovation, and Growth," *NBER Working Paper No. 27075*, 2021.

Chellaraj, Gnanaraj, Keith Maskus, and Aaditya Mattoo, "The Contribution of International Graduate Students to US Innovation," *Review of International Economics*, 2008, *16*, 444–62.

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.*

Doran, Kirk, Alexander Gelber, and Adam Isen, "The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries," *Journal of Political Economy*, 2022, *130* (10), 2501–2533.

Doran, Kirk Bennett, Alexander Gelber, and Adam Isen, "The Effect of High-Skilled Immigration on Patenting and Employment: Evidence from H-1B Visa Lotteries," *NBER Working Paper No. 20668*, 2014.

Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 2010, *100*, 1195–213.

Fadlon, Itzik and Torben Heien Nielsen, "Family Labor Supply Responses to Severe Health Shocks," *American Economic Journal: Applied Economics*, 2021, *13* (3), 1–30.

Foley, C Fritz and William Kerr, "Ethnic Innovation and US Multinational Firm Activity," *Management Science*, 2013, *59* (7), 1529–44.

Gelbach, Jonah, "When Do Covariates Matter? And Which Ones, and How Much?," *Journal of Labor Economics*, 2016, *34* (2), 509–43.

Hall, Bronwyn, 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*, 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.

_ , Jean-Philippe Garant, Hannah Herman, and David Munroe, "Why are Women Underrepresented Amongst Patentees?," *Research Policy*, 2013, *42* (4), 831–43.

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

Jaffe, Adam, "Real Effects of Academic Research," *American economic review*, 1989, *79* (5), 957–70.

_ 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, "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?," *Review of Economic Studies*, 2009, *76* (1), 283–317.

_ , “Age and Great Invention,” *Review of Economics and Statistics*, 2010, *92* (1), 1–14.

_ **and Benjamin Olken**, “Do Leaders Matter? National Leadership and Growth Since World War II,” *Quarterly Journal of Economics*, 2005, *120* (3), 835–864.

_ **and Bruce Weinberg**, “Age Dynamics in Scientific Creativity,” *Proceedings of the National Academy of Sciences*, 2011, *108* (47), 18910–14.

_ **, EJ Reedy, and Bruce Weinberg**, “Age and Scientific Genius,” Technical Report, National Bureau of Economic Research 2014.

Jones, Charles, “R & D-based Models of Economic Growth,” *Journal of political Economy*, 1995, *103* (4), 759–84.

Kerr, William, “The Ethnic Composition of US Inventors,” *HBS Working Paper No. 08-006*, 2008.

_ , “Ethnic Scientific Communities and International Technology Diffusion,” *Review of Economics and Statistics*, 2008, *90* (3), 518–37.

_ , “The Agglomeration of US Ethnic Inventors,” *NBER Working Paper No. 15501*, 2010.

_ **and William Lincoln**, “The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention,” *Journal of Labor Economics*, 2010, *28* (3), 473–508.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar, “Who Profits From Patents? Rent-Sharing at Innovative Firms,” *Quarterly Journal of Economics*, 2019, *134* (3), 1343–404.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, “Technological Innovation, Resource Allocation, and Growth,” *Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Lanjouw, Jean, 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, *Age and Achievement American Philosophical Society*, Princeton University Press, 1953.

Lerner, Josh, Morten Sorensen, and Per Strömberg, “Private Equity and Long-run Investment: The Case of Innovation,” *Journal of Finance*, 2011, *66* (2), 445–77.

Levin, Sharon and Paula Stephan, “Research Productivity Over the Life Cycle: Evidence for Academic Scientists,” *American Economic Review*, 1991, *81* (1), 114–32.

Marshall, Alfred, *Some Aspects of Competition*, Harrison and Sons, 1890.

McDowell, John, “Obsolescence of Knowledge and Career Publication Profiles: Some Evidence of Differences among Fields in Costs of Interrupted Careers,” *American Economic Review*, 1982, *72* (4).

Moser, Petra, “Taste-based Discrimination Evidence from a Shift in Ethnic Preferences after WWI,” *Explorations in Economic History*, 2012, *49* (2), 167–88.

—, **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), 550–67.

Oettl, Alexander, "Reconceptualizing Stars: Scientist Helpfulness and Peer Performance," *Management Science*, 2012, *58* (6), 1122–40.

Oster, Sharon M and Daniel S Hamermesh, "Aging and Productivity Among Economists," *Review of Economics and Statistics*, 1998, *80* (1), 154–6.

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

Romer, Paul, "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5, Part 2), S71–S102.

Roth, Jonathan, Pedro Sant'Anna, Alyssa Bilinski, and John Poe, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *arXiv:2201.01194v3*, 2022.

Saxenian, Anna Lee, "Silicon Valley's New Immigrant High-growth Entrepreneurs," *Economic Development Quarterly*, 2002, *16* (1), 20–31.

Saxenian, AnnaLee, Yasuyuki Motoyama, and Xiaohong Quan, *Local and Global Networks of Immigrant Professionals in Silicon Valley*, Public Policy Institute of CA, 2002.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian, "Immigrants and the Making of America," *The Review of Economic Studies*, 2020, *87* (1), 382–419.

Seru, Amit, "Firm Boundaries Matter: Evidence from Conglomerates and R&D Activity," *Journal of Financial Economics*, 2014, *111* (2), 381–405.

Simonton, Dean, "Creative Productivity Through the Adult Years," *Generations: Journal of the American Society on Aging*, 1991, *15* (2), 13–6.

—, "Emergence and Realization of Genius: The Lives and Works of 120 Classical Composers," *Journal of Personality and Social Psychology*, 1991, *61* (5), 82940.

Stephan, Paula and Sharon Levin, "Age and the Nobel Prize Revisited," *Scientometrics*, 1993, *28* (3), 387–99.

Usher, Abbott Payson, *A History of Mechanical Innovations*, Cambridge, Mass, 1954.

Weitzman, Martin, "Recombinant Growth," *Quarterly Journal of Economics*, 1998, *113* (2), 331–60.

Yonker, Scott E, "Geography and the market for CEOs," *Management Science*, 2017, *63* (3), 609–630.

Zuckerman, Harriet, *Scientific Elite: Nobel Laureates in the United States*, Transaction Publishers, 1977.

Appendix

A Matching Algorithm of Patent Data with Infutor

In this section, we discuss how we build our inventor identifiers in the patent data. The idea is to use individuals' address histories in Infutor data to verify the address at which each inventor lived and disambiguate inventors who moved. Precisely, in the patent data, we define "identifiers" (ID) as a unique combination of first name, last name, city name, and the state in which the inventor lived. In the raw USPTO data, we start with 14,991,282 patents granted worldwide and restrict to 7,350,977 patents granted in the U.S and then to 7,228,174 patents for which state and city name are not missing. Since Infutor data span 1990 to 2016, we necessarily restrict the patent data to these years only. At the start, there are 6,229,618 patents granted to 1,351,024 unique IDs. We discuss our disambiguation steps in detail as follows.

We begin by matching each observation in the patent data to address histories in Infutor data. Our matching criterion is such that state, city, last name, and the first three letters of first name exactly match in the two datasets: 1,034,288 unique IDs have at least one Infutor match in this stage. In this matched subset, each ID can still map to multiple individuals in Infutor. Thus, we successively apply disambiguation restrictions, in which we identify unique matches and then remove them from the pool before applying the next restriction.

First, we impose that first name matches exactly and that middle initial does not conflict between the two datasets (i.e., they agree or at least one is missing). Now, given typographical inconsistencies in first names observed between the two datasets, we allow first names to match weakly. Second, we impose that first name can contain each other (e.g., "Timothy" and "Tim") and that middle initial does not conflict. Third, we allow for alternate first names (e.g., "Richard" and "Rick") and for minor misspelling (e.g., "Stephen" and "Steven"'), while maintaining that middle initial does not conflict. In these steps, we identify 876,438 IDs, each of which maps to exactly one individual in Infutor: 680,261, 144,431, and 51,746 IDs from the first, second, and third steps, respectively.

Next, we try to disambiguate the remaining 157,850 IDs for which state, city, last name, and the first three letters of first name match exactly but for which we cannot find unique Infutor matches in the first three steps. In each of the following steps, we always condition on observing at least one patent of which application year falls between beginning and ending address years (allowing plus or minus two years). In addition to this condition, in each step, we include a successively stricter matching criterion. IDs are identified at the end of each step if we only find one unique Infutor match. For brevity, we define the following terms. Middle initial "matches strictly" if it is nonmissing and agrees across the two datasets and " matches weakly" if it is missing in at least one dataset. First name "matches strictly" if it agrees across the two datasets and "matches weakly" if one contains the other, is an alternate name for the other, or contains a minor misspelling by at most two characters.

In the fourth step, we require weak middle initial match and weak first name match. The fifth step further imposes strict first name match in addition to weak middle initial match. In the sixth step, we condition on strict middle name match and weak first name match. The seventh step then imposes strict first name match in addition to strict middle name match. The eighth step considers cases when both middle name (rather than its first letter) and first name match exactly across the two datasets. We identify 51,622 IDs with unique Infutor matches in these steps: 30,709, 14,457, 2,443, 652, and 3,401, respectively. Altogether, the first eight disambiguation steps identify 928,100 IDs.

Finally, we turn to the 316,736 IDs in the patent data for which we cannot find matches in Infutor using exact matches on state, city, last name, and the first three letters of the first name. To do so, we match these observations to Infutor using exact matches on state, city, and last name, ignoring first name. In essence, these observations are those with inconsistent first three letters of first name. Then, we condition our matches on a strict middle initial match and a weak first name match (as defined above). This final step yields 3,529 IDs.

In summary, out of the 1,351,024 IDs in the patent data, we find 931,629 IDs with unique Infutor matches, indicating a match rate of 69.0% and corresponding to 879,988 unique inventors/individuals in Infutor. We summarize all of these data construction steps in Appendix Table A.2 below

Figure A.1: Validation with CENSUS 2000 - Population Sizes (millions)

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

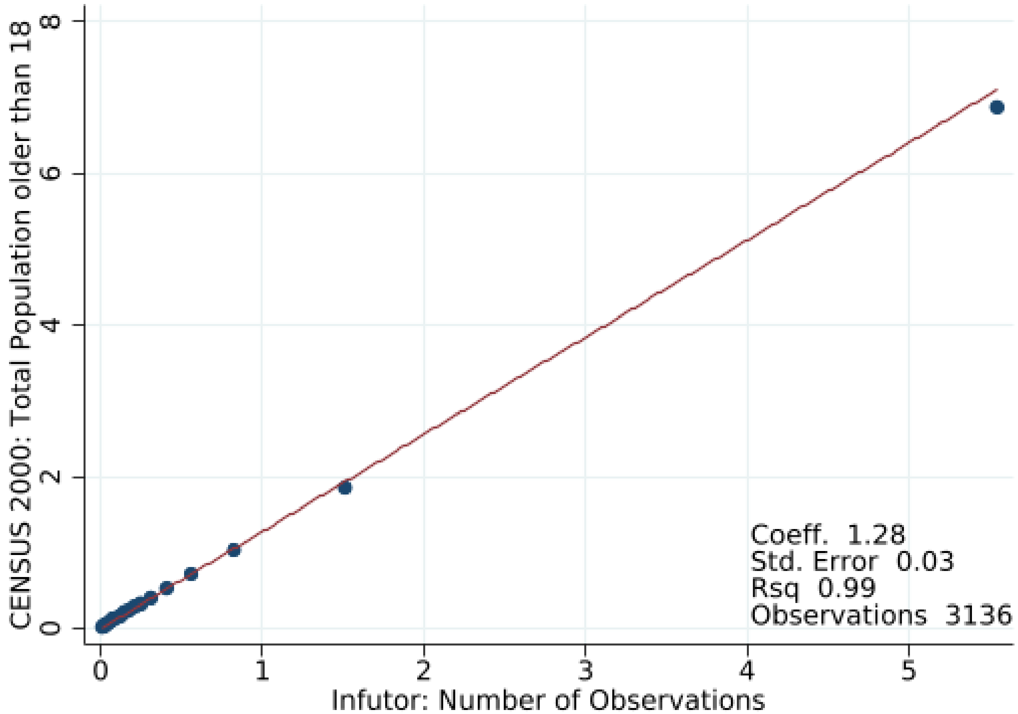


Figure A.2: Validation of Pre-1950 Assignment Year Imputation

Notes. Binscatter of the encoded group numbers for each assignment year, constructed after controlling for fixed effects of area code and weighted by the number of observations in each area and group. Assignment year was collected 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.

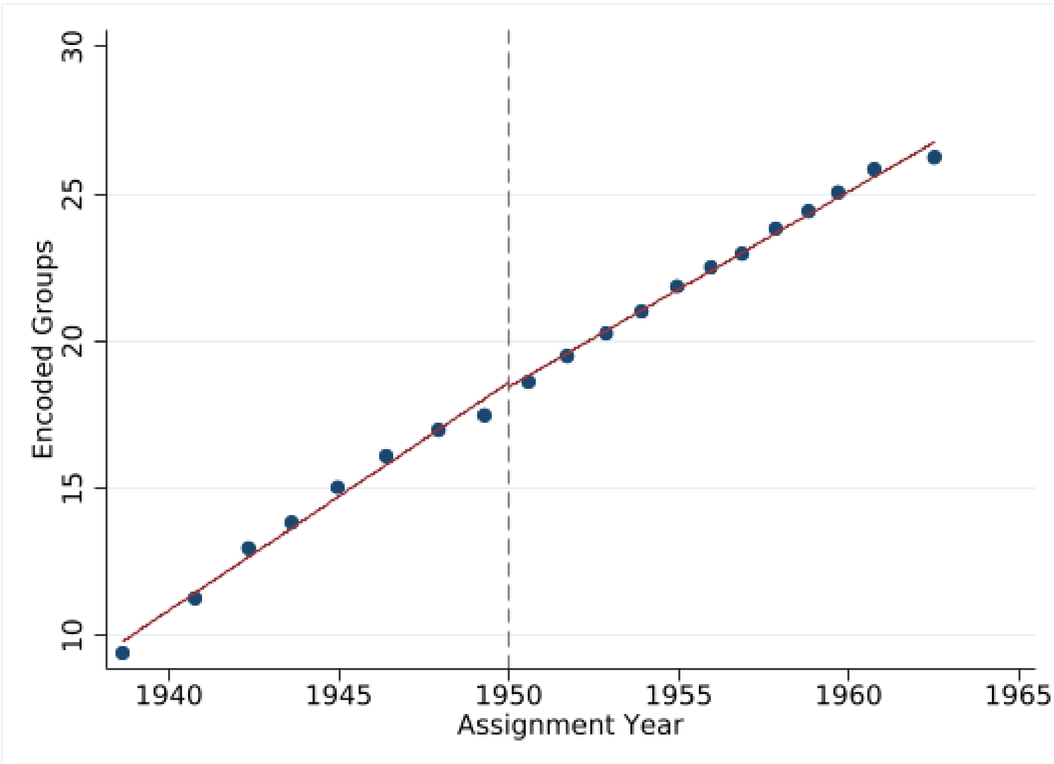


Figure A.3: SSN Issuance Age Distribution

Notes. Quantiles of the age of SSN issuance distribution by assignment year, calculated at the individual level. Assignment year was collected 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.

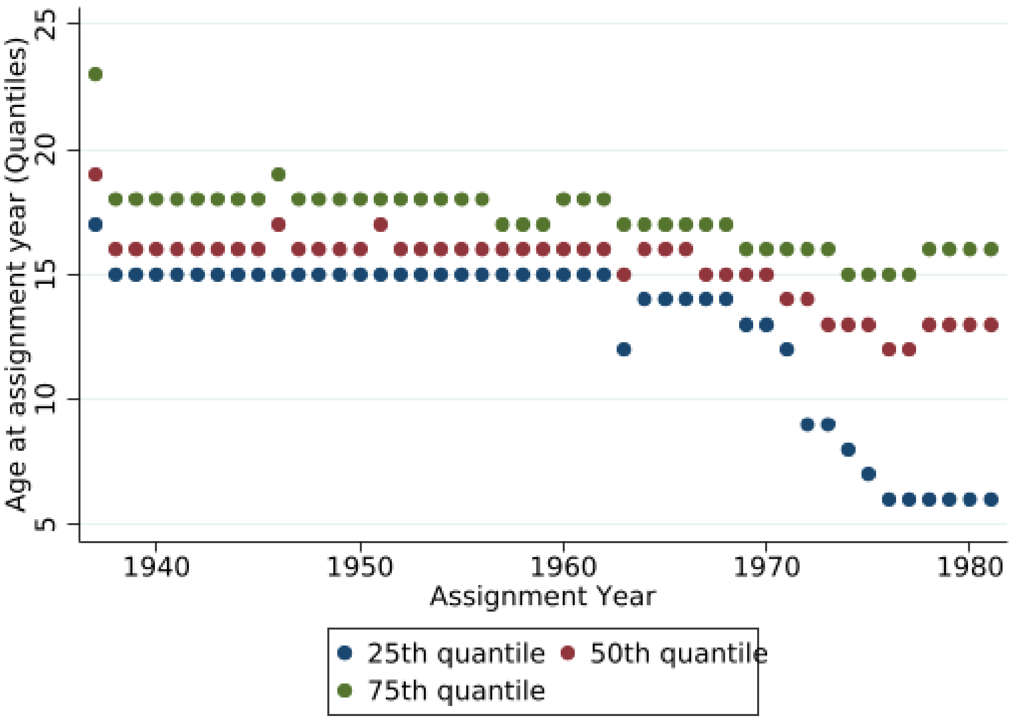


Figure A.4: Validation with CENSUS 2000

Notes. R^2 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 places 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 are weighted by the total population at that county in CENSUS 2000.

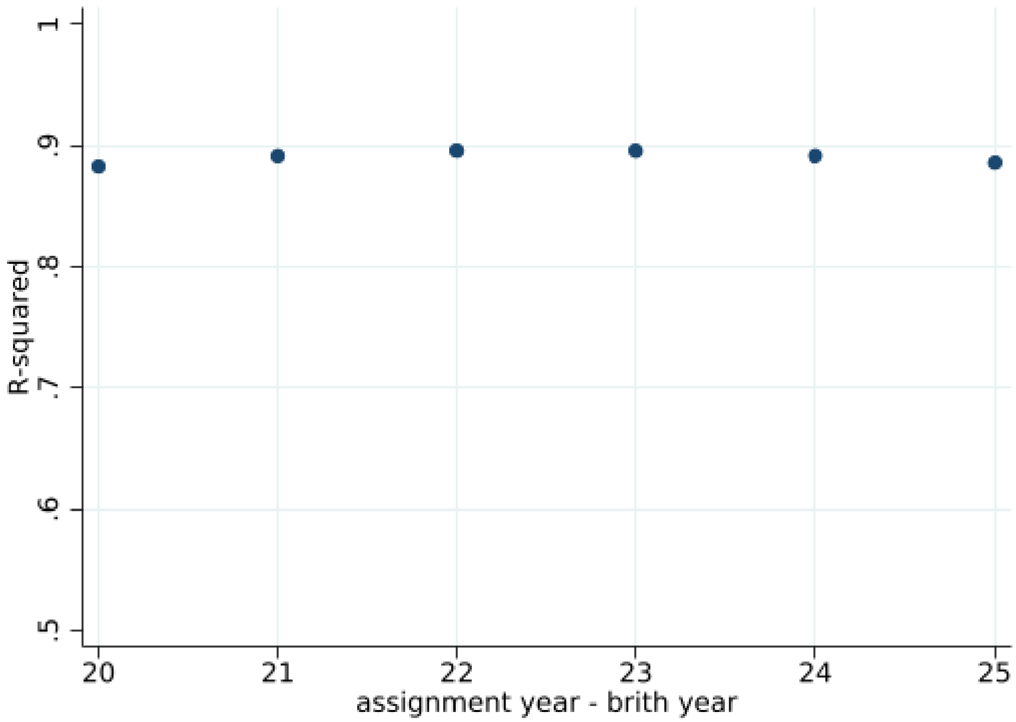
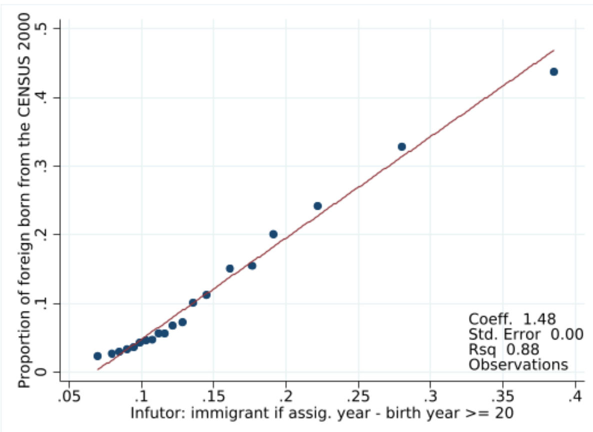


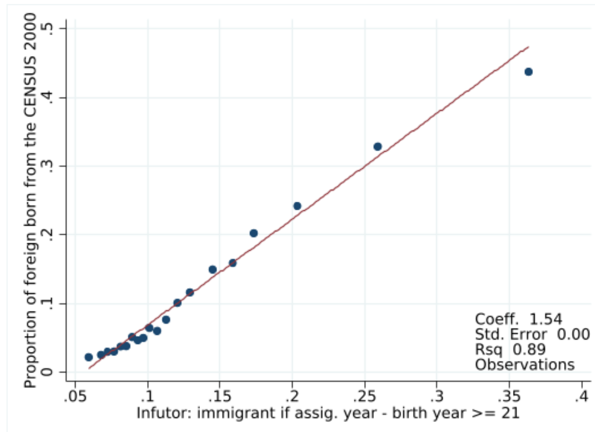
Figure A.5: Validation with CENSUS 2000 – Binscatters

Notes. Binscatters of the proportion of foreign born in the CENSUS 2000 against the proportion of immigrants among all individuals that Infutor places 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 are weighted by the total population at that county in the CENSUS 2000.

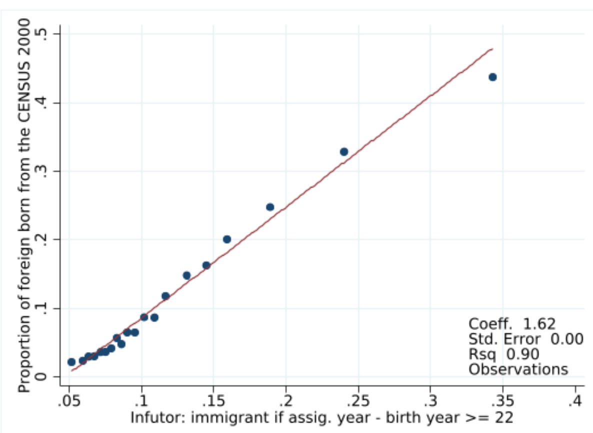
(a) **Immigrant if assig. year − birth year ≥ 20**



(b) **Immigrant if assig. year − birth year ≥ 21**



(c) **Immigrant if assig. year − birth year ≥ 22**



(d) **Immigrant if assig. year − birth year ≥ 23**

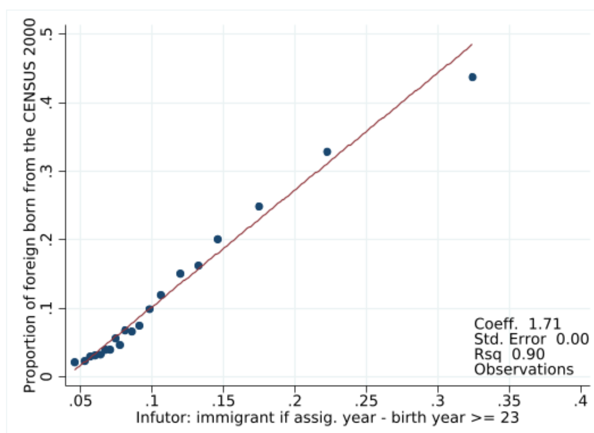


Figure A.6: Validation with CENSUS 2000

Notes. R^2 and slope coefficient of regressing at the county level the proportion of foreign born in CENSUS 2000 against the proportion of immigrants among all individuals that Infutor places 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 are weighted by the total population at that county in CENSUS 2000.

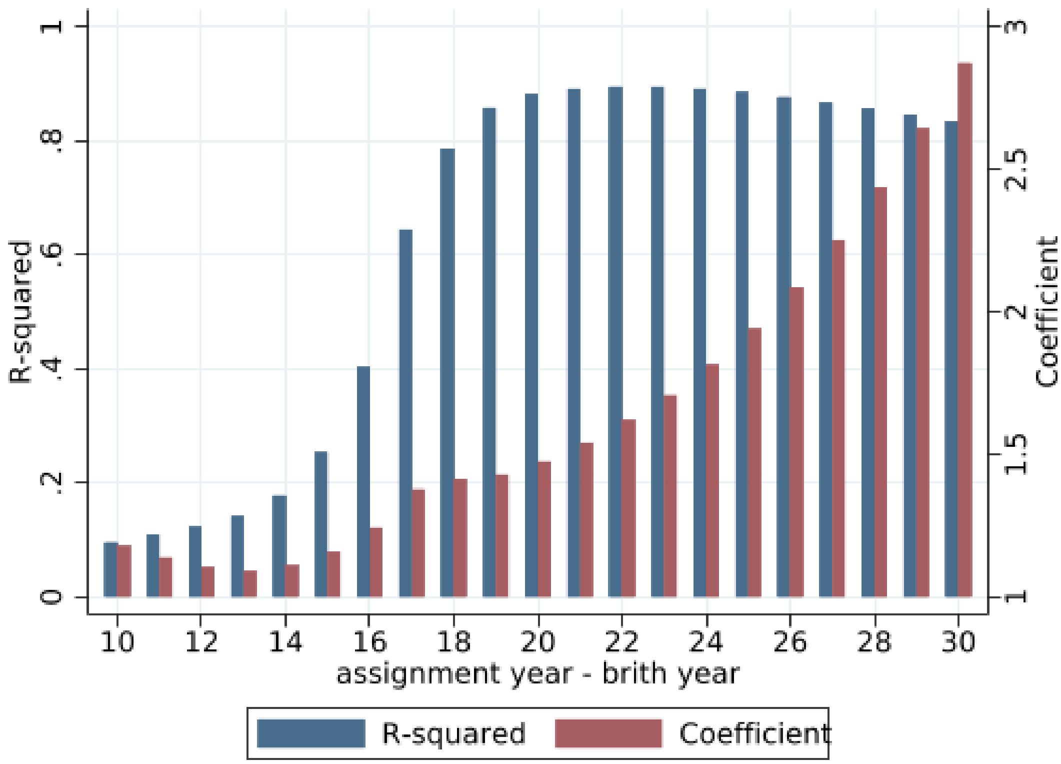
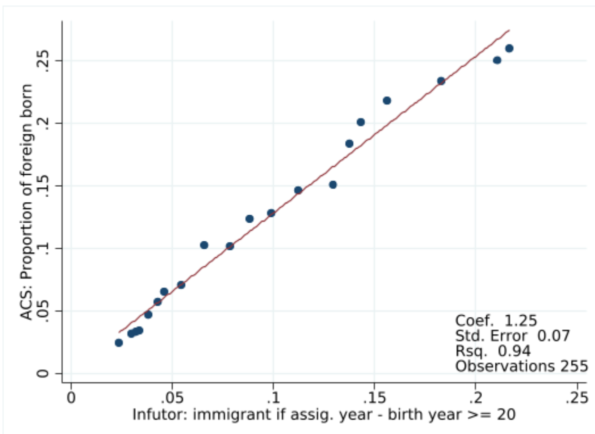


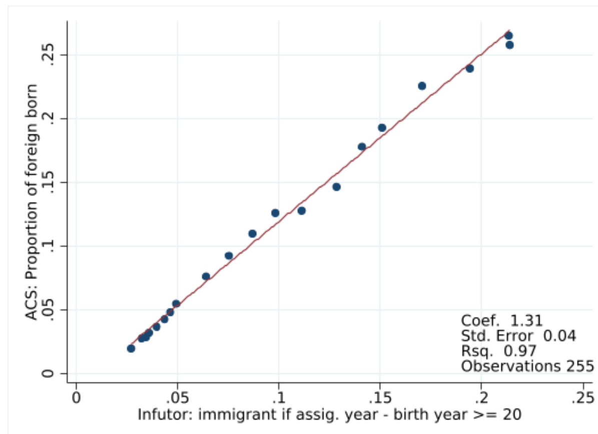
Figure A.7: Validation against ACS by Selected Age Bins in 2005

Notes. 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 separate regression. All regressions are 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.

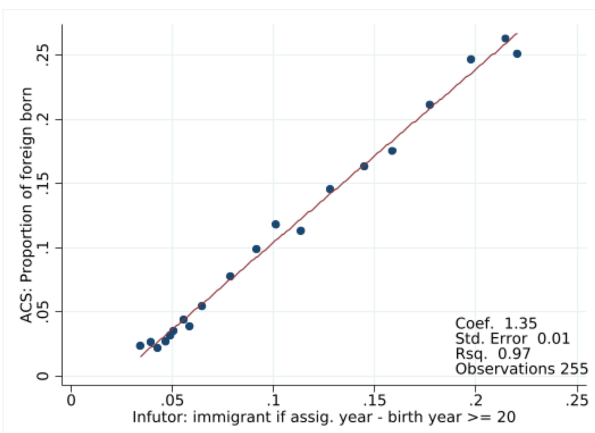
(a) **40-44 years old**



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



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



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

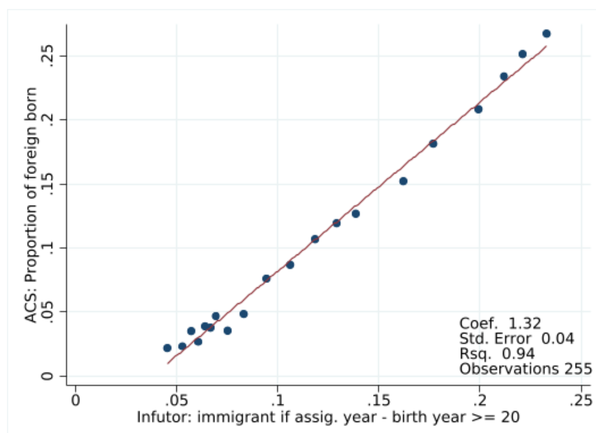
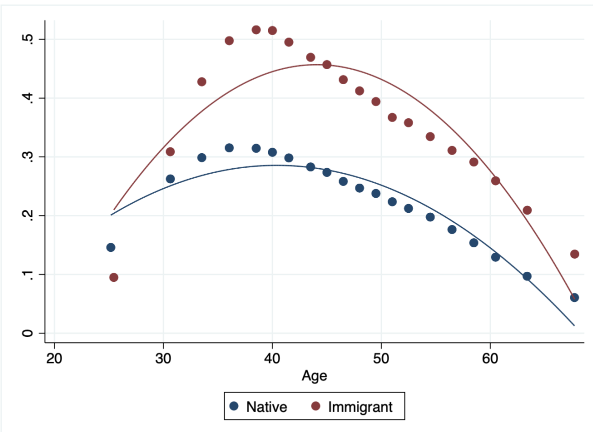


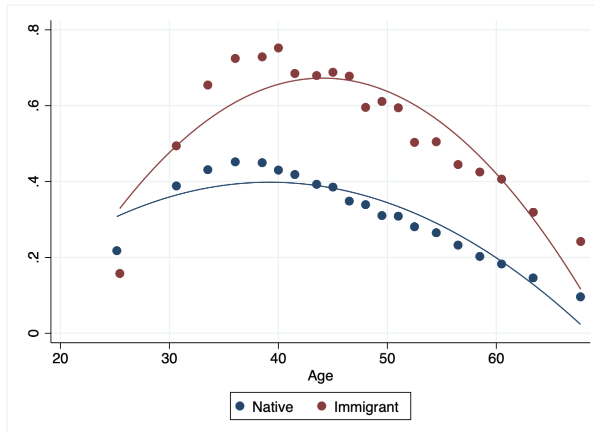
Figure A.8: Productivity over the Life Cycle - First Patent in 1990s

Notes. 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; and (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 and imputed for private firms. Only individuals who applied for their first patent between 1990 and 1999.

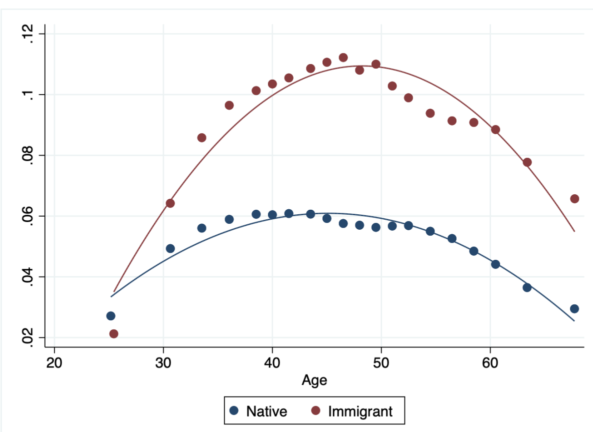
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

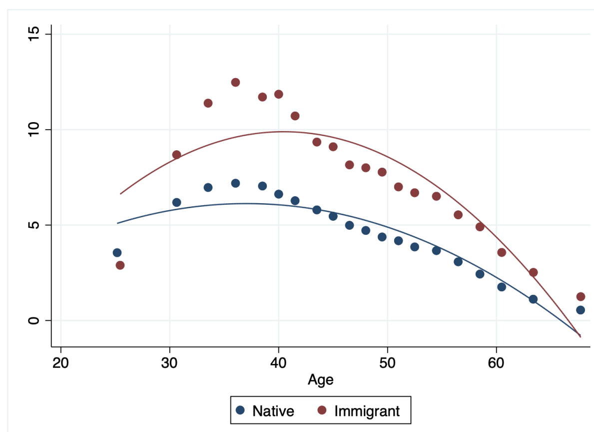
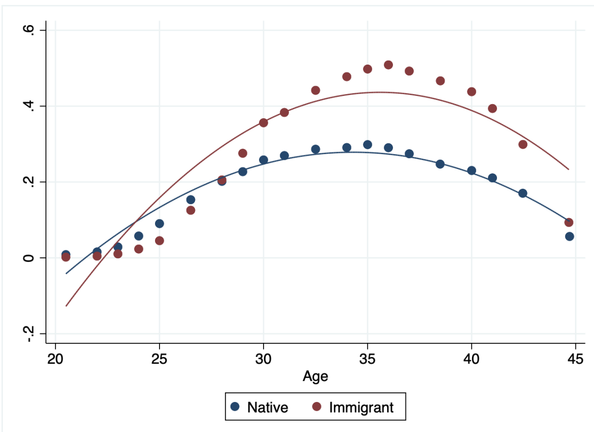


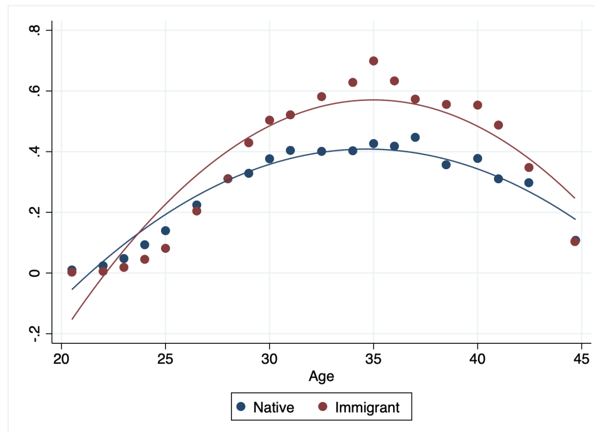
Figure A.9: Productivity over the Life Cycle - 1970s Year of Birth

Notes. 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; and (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 and imputed for private firms. Only individuals born between 1970 and 1979.

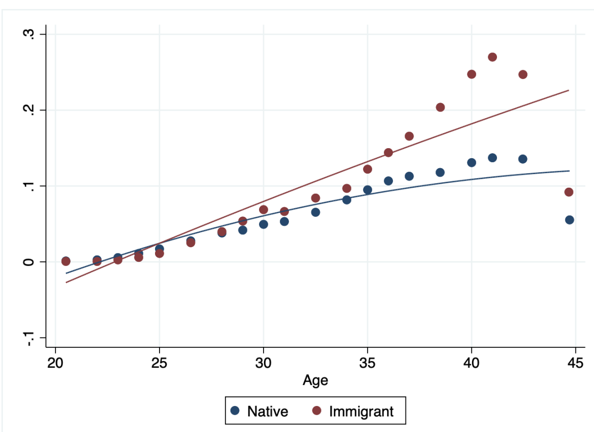
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

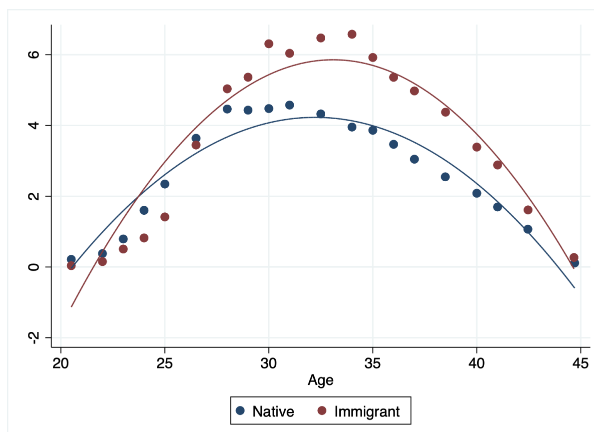
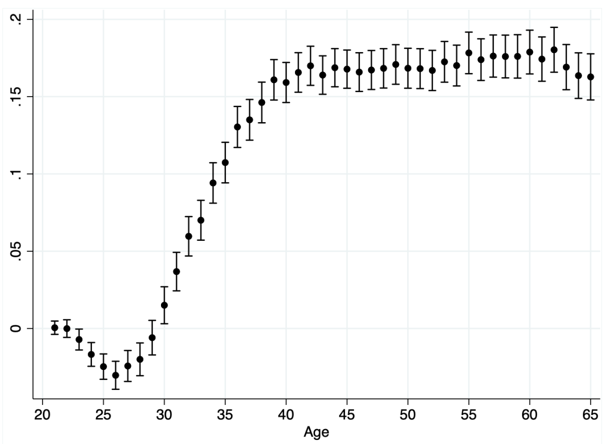


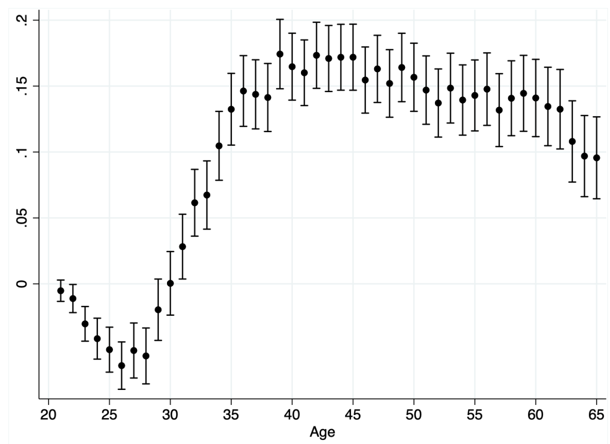
Figure A.10: Productivity over the Life Cycle - Regressions

Notes. 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; and (d) Patent value calculated based on stock market reaction to patent approval using the KPSS measure which is available for publicly traded firms and imputed for private firms.

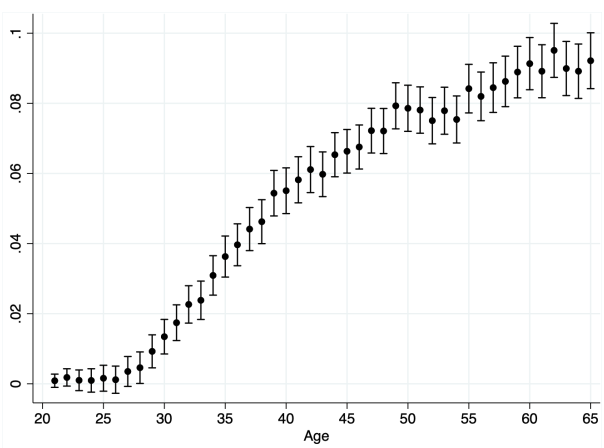
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

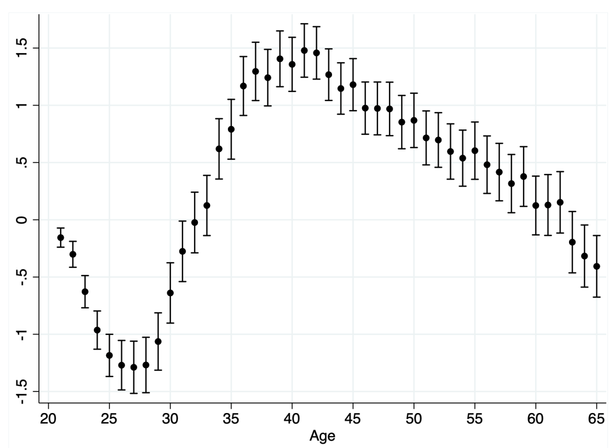


Table A.1: **Prediction of KPSS Economic Value**

Notes. This table reports the relationship between KPSS economic value and patent application and assignee-level characteristics, following a similar imputation in Kline et al. (2019). Coefficient estimates are based on a Poisson model with technology class random effects. The sample is the subsample of granted patents for which the Kogan et al. (2017) measure of economic value is available in our analysis sample. The dependent variable is the KPSS measure of economics value in millions of dollars. Standard errors are reported in parentheses. Number of claims measures the number of claims in the published U.S. patent application. $\log(\sigma_v)$ reports the log of the estimated standard deviation of the technology class random effects. χ^2 reports a likelihood ratio test statistic against a restricted Poisson model without random effects.

	KPSS Value	
1(number of claims = 1)	0.2737***	(0.0025)
log(number of claims)	0.1793***	(0.0003)
Application year	0.0026***	(0.0002)
(Application year)^2	0.0035***	(0.0000)
Decision/grant year	0.0267***	(0.0002)
(Decision/grant year)^2	-0.0099***	(0.0000)
Constant	2.2346***	(0.0373)
$\log(\sigma_v)$	0.7559***	(0.0522)
Technology class	573	
χ^2	2.67×10^6	
N	1,425,642	

Table A.2: **Sample Construction**

Data Processing Steps	# Patents	# Unique IDs
Panel A. Cleaning Raw Patent Data		
– Start with raw USPTO data	14,991,282	
– Keep only U.S. patents	7,350,977	
– Keep if state and city are not missing	7,228,174	
– Keep if application year is between 1990 and 2016	6,229,618	1,351,024
	(100%)	(100%)
Panel B. Disambiguating Inventors using Infutor Data		
– State, city, last name, and first3 match *and*:		
⋆ Step 1: Middle1 consistent, exact first name match	3,287,142	680,261
⋆ Step 2: Middle1 consistent, contained first name match	673,854	144,431
⋆ Step 3: Middle1 consistent, alternate or misspelled first name	237,993	51,746
– At least one patent with consistent application-address year *and*:		
⋆ Step 4: Weak middle1 match, weak first name match	79,859	30,709
⋆ Step 5: Weak middle1 match, strict first name match	84,531	14,457
⋆ Step 6: Strict middle1 match, weak first name match	7,869	2,443
⋆ Step 7: Strict middle1 match, strict first name match	3,388	652
⋆ Step 8: Both middle name and first name match exactly	17,589	3,401
– State, city, last name match *and*:		
⋆ Step 9: Middle1 consistent, exact/contained/alternate/misspelled first name	14,986	3,529
– Number of IDs with unique Infutor matches	4,407,211	931,629
	(70.7%)	(69.0%)
– Number of unique inventors/individuals in Infutor		879,988

Notes: Identifier (ID) is a combination of state, city, last name, and first name in the patent data. "Middle1" stands for the first letter of middle name. "Weak middle1 match" is when middle1 is missing in at least one dataset (USPTO or Infutor). "Strict middle1 match" is when middle1 is non-missing and the same across the two datasets. "Middle 1 consistent" is either weak or strict middle1 match. "First3" stands for the first three letters of first name. "Weak first name match" is when first name is either contained within or an alternate name for each other or contains misspelling. "Strict first name match" is when first name matches exactly across the two datasets.