NBER WORKING PAPER SERIES

THE CONTRIBUTION OF HIGH-SKILLED IMMIGRANTS TO INNOVATION IN THE UNITED STATES

Shai Bernstein
Rebecca Diamond
Abhisit Jiranaphawiboon
Timothy McQuade
Beatriz Pousada

Working Paper 30797
http://www.nber.org/papers/w30797

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2022, Revised August 2025

Shai Bernstein is with Harvard University and NBER, Rebecca Diamond is with Stanford University and NBER, Timothy McQuade is with UC Berkeley and NBER, Abhisit Jiranaphawiboon is with Stanford University, and Beatriz Pousada is with Charles River Associates. We thank seminar and conference participants for helpful comments and feedback. The authors have obtained IRB approval from Stanford University before conducting the analysis. Research support was provided by Stanford GSB research funds. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Contribution of High-Skilled Immigrants to Innovation in the United States
Shai Bernstein, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada
NBER Working Paper No. 30797
December 2022, Revised August 2025
JEL No. J6, O31

ABSTRACT

We characterize the contribution of immigrants to US innovation. Leveraging new data, we use age of SSN assignment to identify immigrant status. Immigrants represent 16 percent of inventors, but authored 23 percent of patents. Immigrant inventors contribute to knowledge diffusion across borders. They disproportionately rely on foreign technologies and inventor collaborations. Using variation from premature inventor deaths, we find immigrant inventors create stronger innovation productivity spillovers on their collaborators, as compared to US-born inventors. A simple model implies immigrants are responsible for 32 percent of aggregate innovation, over half of which is due to human capital externalities on US-born collaborators.

Shai Bernstein
Harvard University
Harvard Business School
and NBER
sbernstein@hbs.edu

Rebecca Diamond
Harvard University
Department of Economics
and NBER
rdiamond@fas.harvard.edu

Abhisit Jiranaphawiboon
Stanford University
ajirana@stanford.edu

Timothy McQuade
University of California, Berkeley
and NBER
tmcquade@haas.berkeley.edu

Beatriz Pousada
Stanford University
bpousada@stanford.edu

1 Introduction

Innovation and technological progress are considered to be key determinants 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] and account for 26% of US-based Nobel Prize winners from 1990 through 2000. A large and growing literature has found that patent data provides a useful laboratory to study the causes and consequences of innovative activity. Using survey data from 2003, Hunt and Gauthier-Loiselle (2010) found immigrants authored 24% of US patents.

In this paper, we characterize the contributions of immigrants to innovation. Clearly, there are many ways to measure innovation, but this paper will focus on the production and impact of US patents as a metric of innovation. We bring to bear administrative patent data and a unique approach to identifying the immigrant status of individuals residing in the United States at scale. First, we use the richness of our data to describe the life-cycle patterns of patenting by immigrant status and investigate possible mechanisms driving productivity differences between immigrants and the US born. Second, we quantify the extent to which inventors' patenting productivity is impacted by collaborations with other inventors, and how these productivity spillovers vary according to immigrant status. Finally, using these estimates, we quantify the share of aggregate US patenting production that can be attributed to immigrants, inclusive of their indirect productivity spillovers on US-born inventors.

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. We link the universe of patents from 1990-2016 to the Infutor data based on a merge of first and last name, city and state of residence as of the date of the patent. By using the Infutor data as a "back bone" for the patent data, we can disambiguate which inventors are the same person across patents. This provides an alternative disambiguation procedure to Balsmeier et al. (2015).[2] 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-born individuals 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.[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. In our sample from

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

[2]A benefit of our disambiguation method is that we do not need to rely on similarity of patenting technology class across patents or assumptions of geographic immobility of inventors across patents. Indeed, when studying how inventors choose their topics of innovation and geographic locations, using an identification method that relies on these not changing over time can lead to biased estimates.

[3]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.

1990 to 2016, we find immigrants that came to the United States when they were 20 years old or older make up 11% of the US population and 16% of all US-based inventors. 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 and more than a 100% increase relative to their share of the total US population. This finding is consistent with Hunt and Gauthier-Loiselle (2010) who find that 24% of patents in 2003 are authored by immigrants, while constituting 14% of the inventor population, according to survey evidence from the National Survey of College Graduates (NSCG).[4,5] Relative to this prior work, by having access to administrative patent data, we are able to observe granular information regarding patent characteristics and patent quality by immigrant status, as well as panel information of patenting activity over the life-cycle.

In particular, immigrant patents do not appear to be of lower impact. Using the number of patents weighted by the number of forward citations, which captures the impact of innovation (Hall et al., 2001), we find that the immigrant contribution is slightly 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.

Descriptive trends of inventor life-cycle productivity show both US-born and immigrant inventors exhibit an inverse U-shape pattern. Inventors are quite unproductive at the beginning of their careers, become most productive in their late 30s and early 40s, and decline in productivity thereafter.[6] However, immigrant inventors diverge from US-born inventors as they reach 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.[7]

Prior descriptive work using administrative patent data in the modern period has focused on documenting time trends in patenting output across inventors of different ethnic origins, with ethnicity imputed according to their first and last name (Kerr, 2008b,a, 2010; Foley and Kerr,

[4] Key advantages of the National Survey of College Graduates are that it allows for identification the type of visa used by immigrants. 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. Moreover, Hunt and Gauthier-Loiselle (2010) describe the role of occupation and education in inventor productivity by immigrant status, which we are unable to observe.

[5] Doran et al. (2014) exploit random variation H1-B lottery winners to argue that additional H1-B workers lead to only modest, insignificant impacts on total innovation at the firm level. However, this estimate is a local-average treatment effect (LATE), analyzing the incremental patenting generated by the marginal immigrant selected by H1-B lotteries. In contrast, our work studies patenting output of the average US-based immigrant patenting between 1990 and 2016. Indeed many immigrants are not on H1B visas. They could have green cards, become US citizens, have O1, J1, or OTP visas. Also, Kerr and Lincoln (2010) exploit large changes in the H1-B program, which likely generates a different complier population, and document increased patenting in exposed firms and cities.

[6] 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.

[7] Akcigit et al. (2017) use the now public-use 1880-1940 censuses to show immigrants were more productive historically as well, producing 9% more patents over their careers than their US-born counterparts. Relative to this work, our paper focuses on the modern period, documents the full patenting life-cycle, addresses differential sorting across location and technology, provides stylized facts on immigrant networks / collaboration, and estimates productivity spillover effects.

2013). For example, Kerr (2008b) uses this approach to show that Chinese and Indian contributions to US technology formation increased dramatically during the 1990s and that ethnic innovation is concentrated in high-tech sectors. While this approach allows one to infer inventor ethnicity, it cannot differentiate foreign-born individuals from US-born individuals as our methodology allows us to do.

While the goal of this paper is not to fully decompose all the reasons immigrant inventors are more productive than US-born inventors over their life-cycle, 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, immigrants disproportionately patent in technology classes that are experiencing more innovation activity. We further document that immigrants disproportionately choose to live in highly productive counties ("innovation hubs"), relative to US born inventors. This latter finding is analogous to Kerr (2010), who shows ethnic inventors are more spatially concentrated than US-born inventors. These findings suggest that immigrant inventors are likely to benefit more from, and contribute more to, agglomeration forces than their US-born counterparts. Ultimately, a flexible regression-based analysis which includes county by technology class by year fixed effects, as well as year-of-birth cohort fixed effects, can explain approximately 30% of the raw patenting gap between immigrant and US-born inventors.

We also provide suggestive reduced-form evidence 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 foreign citations. Immigrants are also about twice as likely to collaborate with foreign inventors, relative to US-born inventors. Finally, foreign inventors are about ten percentage points more likely to cite the patents of US-based immigrant inventors relative to patents of US-born inventors. These results complement those of Foley and Kerr (2013), who show that increases in a multinational firm's patenting activity by a particular ethnicity are associated with increases of affiliate activity in countries related to that ethnicity.

We also address the extent to which immigrant inventors are integrated in US knowledge markets and the extent to which they collaborate with their US-born inventor counterparts. One possibility is that, due to cultural impediments or lack of assimilation, immigrant inventors are less integrated into the overall US knowledge market, remain isolated at their workplace, and collaborate less, which Jaravel et al. (2018) document is important to the innovative process. We show this is not the case. Immigrant inventors in fact tend to have more collaborators than US-born inventors. Furthermore, while we do find that immigrants are more likely to work with other immigrants (as compared to US-born), this tendency declines over the life-cycle, suggesting a gradual assimilation process.

These team interactions between immigrant 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 next turn to the secondary goal of the paper of measuring the extent to which immigrant inventors, through collaboration, make other US-based inventors more productive in their patenting.

We estimate the magnitudes of immigrant 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.[8] 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 US-born inventor deaths. The death of an immigrant lowers co-inventor productivity by approximately 16%, 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. For example, if more productive inventors have larger spillover effects, and dying immigrant inventors are more productive, then controlling for productivity-driven spillover effects could narrow the differential spillover effects between immigrant and US-born inventors. Even after controlling for a host of observable characteristics interacted with treatment, which could explain the heterogeneous treatment effects between US-born and immigrant inventors, the productivity spillover gap between US-born inventors and immigrant inventors remains essentially unchanged. To further understand why the gap remains stable, we estimate a Gelbach (2016) decomposition of the difference based on ten dimensions of treatment effect heterogeneity. We find 15% of the immigrant-US-born productivity spillover gap can be explained by dying inventors who have more patents at time of death have larger effects on their surviving collaborators. However, the immigrant-US-born productivity gap is *widened* by 54% due to surviving collaborators with more prior patents being less impacted by a dying inventor, and immigrant collaborators having more patents than US-born collaborators. Thus, controlling for these observable differences, along with measures of age, cohort, measures of the collaboration network, recency, and intensity, knowledge overlap between collaborators, and geography does little to explain away the larger external effect of immigrant inventor deaths. Our inability to reduce

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

the gap, despite controlling for many observable differences, provides evidence that there is something unique about immigrant inventors that drives large productivity spillovers on their US-based co-authors, which cannot be easily replicated.

Given the nature of our inventor-patent panel, our work is unable to speak to possible crowd-out effects, which could exist in conjunction with positive productivity spillover effects. Specifically, we are unable to track whether individuals with no prior relationship to the dying inventor now start to patent *more*, replacing some of the knowledge that would have been produced by the dying inventor and his/her team. Prior work on the extent of crowd-out is mixed. Borjas and Doran (2012) provide evidence that the post-1992 influx of Soviet mathematicians crowded out young American scholars, with total US mathematical output remaining stable. In contrast, Moser et al. (2014) show that the influx of Jewish immigrant chemists from Nazi Germany increased patenting in chemistry by 71%, driven by the entrance of new domestic scientists. Doran et al. (2014) use H1-B lottery data to argue that additional H1-B workers crowd out other workers at the firm and lead to only modest, insignificant impacts on total firm innovation. Conversely, using a different firm sample, Brinatti et al. (2023) find no evidence that winning H1-B lotteries leads to net displacement of US-born workers. Similarly, Kerr and Lincoln (2010) exploit large changes in the H1-B program and find increases in patenting in exposed firms and cities, with evidence of small crowd-in effects as opposed to crowd-out.

Prior work by Hunt and Gauthier-Loiselle (2010) suggests that positive spillover effects of immigrants on patenting exceed any crowd-out effects. They instrument for changes in immigrant college populations using a 1940-2000 state panel and show that a 1 percentage point increase in the immigrant college share increases patents per capita by 9-18 percent, larger than the 6 percent increase suggested by the raw patenting differences.[9] We build on this prior work by exploiting micro information regarding individual inventor teams to cleanly isolate one specific source of spillovers, collaboration externalities. We show that immigrants contribute in a disproportionate way to the productivity of their collaborators, relative to their US-born counterparts.

Our analysis raises the question of whether positive spillover effects of immigrants on their US-born collaborators are confined to the teams they are both a part of, increasing team-specific capital, or if they reflect human capital spillovers that increase the productivity of their collaborators on teams which the immigrant is not a part of. To address this question, we construct and estimate a team-based structural patent production function, which allows us to quantify these two channels separately. We find that immigrants increase team-specific capital, but also increase the transferable human capital of their US-born collaborators.

Finally, we use this structural framework to quantify the share of aggregate innovation which can be attributed to immigrants, both through their direct output and indirect human capital spillovers. We conclude that 32% of total US innovative output, since 1990, can be ascribed

[9] Cristelli and Lissoni (2020) show the Swiss 1999 easing of border restrictions led to increased patenting in the border regions and increased collaboration with cross-border inventors, with no evidence of crowd-out. Wigger (2022) uses exogenous variation in push factors to show high-skilled immigration within Europe led to increased regional patenting, driven in part by increased immigrant-native collaboration and increased native patenting.

to US-based immigrants, despite only making up 16 percent of the inventor workforce and only directly authoring 23% of patents. This additional 9 percentage points of innovation, over and beyond immigrants' direct output, is due to immigrants' substantial human capital externalities on US-born inventors. Moreover, the decomposition also highlights the importance of the two-way spillovers between immigrant inventors and US-born inventors, with one-quarter of US innovation attributable to this source.

The remainder of the paper proceeds as follows. Section 2 describes the data used in the analysis. Section 3 details our 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 US-born inventor productivity. Section 5 analyzes collaboration externalities of immigrant and US-born inventors and Section 6 provides a structural framework to separately quantify immigrant contributions to team-specific capital and transferable human capital, as well as to quantify the aggregate contribution of immigrants to total innovation. 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 US-born inventors in the United States. Specifically, we combine patent data from the US Patent Office (USPTO) together with data provided by Infutor, which allows us 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.[10] 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. (2019).

This data appears to be highly representative of the overall US adult population.[11] 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

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

[11]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.

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 2016 directly from the United States Patent and Trademark Office (USPTO). The USPTO data provides information on the date a patent was applied for and ultimately granted, the individual(s) credited as the patent's inventor(s), the firm to which the patent was originally assigned, and other patents cited as prior work. From this, we can determine how many citations a granted patent receives in the future. The data also provides information on the technology class of the patent, as well as the city and state in which each inventor on the patent lives.[12]

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, city, and state of residence at time of patent application. See Appendix A for details.[13] 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%.[14] 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

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

[13] We provide a summary of our USPTO-Infutor matching steps in Table A.1 in the Appendix.

[14] 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.

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 US-born 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).[15] 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 impact 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 impact and economic importance. Specifically, they find that an extra citation per patent boosts a firm's market value by 3%. Similarly, Kogan et al. (2017) find that patent's economic value is strongly correlated with its impact 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) (KPSS) 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. This allows us to measure the KPSS value for the full set of patents, both from public and private firms. The imputation regression is shown in Table A.2.

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

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 pins 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-born individuals are assigned an SSN during their youth, those individuals who receive an 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 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 to age 20. This number falls to 32% among college graduates and 19% among PhDs. This suggests we will classify some immigrants as US-born, implying our analysis focuses on those who immigrate during adulthood.[16] A second issue is that we will miss illegal immigrants, as they would not have SSN. However, this is likely less of an issue for high skilled immigrants who are inventors, since they would likely be employed in the formal sector.

Since our approach relies closely on the structure and precise assignment method of US Social Security numbers, we start by outlining the relevant history and institutional details of the SSN program. We then detail our exact approach to 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

[16] Note that immigrants classified as US-born are unlikely to affect the characteristics of the US-born group given their particularly small fraction relative to the overall group.

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.[17] 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.[18] 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.[19]

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.[20]

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

[18]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).

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

[20]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 plus 16 within the cohort that has that SSN (that is, reflecting the most common age that individuals get their SSN at the time). 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 the 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).

3.2 Identifying Immigrants

Combining this mapping between the first five digits of the SSN and assignment years with individuals' birth year from Infutor, we can use the age at which they are assigned an SSN to classify US-based individuals as either US-born or immigrant individuals.

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.[21] 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 decline considerably.[22]

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.[23] We also explore alternative, more conservative classifications of immigrants, requiring gaps of 21 to 25 years between the SSN assignment year and the individual's birth year. Our results are robust to these alternative classifications. In Appendix A.1, we explore how representative our classification of immigrants is when compared to three different sources of aggregate statistics on immigrants in the United States. We find our data is representative.[24]

3.3 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 the 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

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

[22] 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.

[23] We classify all individuals that have an SSN that is either an ITIN or belongs to Enumeration at Entry program as immigrants as well. To summarize, if we sum all the special cases that we do not 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.

[24] See Figures A.4, A.5, A.6, and A.7 in the Appendix for these additional validation exercises.

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-born inventors. 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

When looking at the number of patents, patents' citations, and patents' economic value, we find that immigrants' contribution to US innovative output is significantly higher than their share of total US-based inventors. Starting by calculating the share of immigrants among US-based inventors, Figure 1 shows that 16% of US-based inventors immigrated to the United States when they were at least 20 years old. This number is in line with statistics provided by the 2016 ACS that says that 16% of workers in STEM occupations were immigrants who immigrated at age 20 or later.[25]

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.[26]

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

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

[26]Figure A.8 shows that our results are robust to apportioning the full value of the patent to each co-inventor on the team.

US-based inventor population. One possibility, though, is that immigrants might be producing more patents of lower impact than their US-born 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 impact. Yet another concern is that immigrants may select into technologies that have higher citation rates, which could account for these results. However, looking at adjusted citations, which scales 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 finally explore the share of economic value that immigrants have generated over the last four decades.[27] We find that immigrants have generated 25% of the aggregate economic value created by patents in publicly traded and private companies between the years of 1990 and 2016.

While we apportion patenting outcomes to team members on an equal-weighted basis, one might still be concerned the above results are driven by differential team sizes between immigrants and the US-born. This would be the case, for instance, if immigrants work on a larger teams and there are increasing returns to scale. First, we show in Figure A.9 that the distribution of team sizes is quite similar between immigrants and US-born inventors. The average team size for US-born and immigrant inventors are 2.366 and 2.300, respectively. Moreover, Figure 1 shows that the disproportionate contributions of immigrants to US-based innovation holds for solo patents as well. Figure A.10 shows that on the intensive margin, controlling for team size, patents produced by teams with more immigrants have larger scaled citation counts.[28]

We additionally 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 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 seems to be lower at about 15% in more traditional technologies such as the "Mechanical" category, which involves Metal working; Transportation; Engines; and the "Other" category, which includes various technologies related to Heating, Agriculture, Furniture, among others.

[27] 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 Table A.2 in the Appendix.

[28] To deal with non-linearity of productivity by team size and differential sorting into different team sizes between US born and immigrant inventors, we have added Table A.3 in the Appendix, which calculates the shares of immigrant contribution across patenting outcomes *for different team sizes*. Holding team size fixed, patents from teams with a larger immigrant share have more citations. Holding fixed immigrant share, patents from larger teams have more impact. Thus, despite the fact that immigrants tend to be on slightly smaller teams, they are still more productive.

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 the US-based inventor population. In this section, we begin to unpack the source of these differences, exploring the innovative productivity of both immigrant and US-born inventors over the life-cycle. To do so, we study patenting activity throughout the span of each inventor's career. We find that both immigrant and US-born inventors show an inverse U-shape productivity over their life cycle, peaking in their late 30s, but immigrant inventors are significantly more productive than US-born inventors at the peak of their productivity.

Panel (a) of Figure 3 illustrates the life-cycle innovative productivity of US-born and immigrant inventors as measured by the annualized number of patents. This figure plots average outcomes by age, separately for immigrant and US-born inventors. These figures plot simple raw means of patenting output over the life cycle using a balanced panel of observations at the inventor-year level. This does not control for time or cohort effects. 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.[29] While the innovative productivity of US-born and immigrant inventors follow similar trajectories early in the life-cycle, the two populations diverge when reaching the peak of innovative productivity, with immigrant inventors significantly more productive than their US-born counterparts. 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 impact 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 immigrant and US-born inventors, we find an inverse U-shape pattern of inventor productivity, but immigrant inventors become significantly more productive than US-born inventors 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 US-born inventors. 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 economic value generated.

The inverse U-shape productivity of US-born 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 so doing, spend less time in active scientific production. Consequently, skill is increasing

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

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 immigrant and US-born inventors early on in the life-cycle, but do not account for the differences in productivity between immigrant and US-born inventors 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 immigrant and US-born inventors, including cohort effects and differential sorting across industries and space. We find that, although they explain part of the gap between immigrant and US-born inventors, they still cannot account for the majority of the difference.

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 impact, 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 US-born productivity could be driven by differences between immigrant and US-born inventors in the distribution of birth years.

Another concern is that immigrant inventors may simply work in different technology classes than US-born inventors. Then, to the extent that certain technology classes are easier to innovate, have more impactful innovations, or have lower burden of knowledge, we would find differences in the innovative output of immigrants versus US-born inventors over their life-cycles. A related concern is immigrant inventors may be differentially sorted into different regions in the United States. To the extent that immigrants, often thought to be more mobile than US-born individuals, 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 US-born inventors 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. In column (6), we include county by technology class by year fixed effects in addition to YOB fixed effects. There is still a substantial productivity gap between immigrant and US-born inventors. 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. Finally, in column (7) we include controls for team size and the productivity of one's co-inventors. Peer quality is defined as the average number of patents and scaled citations across all co-authors within the team (grouped into ventile bins). It is possible the immigrants collaborate with especially productive inventors and thus immigrants are "getting credit" for the disproportionate contributions of their co-inventors. Controlling for team size and peer quality narrows the immigrant-US-born patenting gap to 0.052, but it is still economically large and highly statistically significant.

These results suggest that differential sorting, particularly regional sorting, can explain some of the productivity gap between immigrant and US-born inventors 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 annual production of top patents, and finally in panel (d) we explore the effect on annual economic value. In all of these measures we find that while the gap seem to decrease, between immigrant and US-born inventors, 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 immigrant and US-born inventors still remain when we add such controls, and at the peak of one's career immigrants still remain significantly more productive. See Figures A.11, A.12, and A.13.

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). Therefore, one potential benefit of immigration to the United States 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.

To explore the extent to 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 immigrant and US-born inventors rely on non-US technologies. To do so, for each patent we calculate the share of backward citations of patents that were issued outside the United States. We present the share of foreign backward citations separately for US-born and immigrant inventors 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 US-born inventors. Specifically, on average, immigrants collaborate with at least one foreign inventor in 16% of their patents, by contrast to 9% of US-born inventors. Appendix Figure A.14 further shows that those immigrant inventors that have more foreign co-authors also have more US-born collaborators. In this way, immigrant inventors are well placed to intermediate the flow of knowledge from foreign markets to the US-born.

Finally, in Panel (c), we provide an additional measure that explores the extent to which immigrants are integrated with global innovation markets by exploring how likely foreign inventors are to cite immigrant patents relative to US-born patents. As expected, we find that immigrants' patents are more likely to be cited by foreign inventors. This illustrates the fact that immigrant innovation not only disproportionately draws from foreign markets, but it 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. Moreover, it is interesting to note that the gap between immigrant and US-born 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.[30] Indeed, work by Freeman and Huang (2015) shows scientists with the same ethnicity collaborate more frequently than would be implied by their population shares. Also, immigrants may face discrimination in local labor markets (Moser, 2012).[31] These forces suggest that immigrants may be more inclined to either work in seclusion or,

[30] See Borjas (2014) for a formalization of this idea.

[31] 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

alternatively, may be less inclined to work with US-born inventors. On the other hand, immigrants may collaborate with each other simply because it is beneficial to work with other highly productive inventors. The extent to which immigrants collaborate with US-born inventors may have important implications for the spillovers and the indirect contribution of immigrants to US innovation. The patent data provides a unique glimpse into the assimilation of immigrants to 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, US-born and immigrant inventors 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-born inventors. 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 US-born inventors. 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 US-born inventors, 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 US-born). However, unlike for US-born inventors, 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 their career, immigrants are still more likely to collaborate with other immigrants.

5 Productivity Spillovers

In this section, we explore the extent to which immigrants generate positive spillovers on their collaborators, and whether such spillovers are larger than those for the US-born. Such spillovers could arise from disproportionate contributions of immigrant inventors to team-specific capital; that is, immigrant inventors may increase the productivity of the specific teams on which they work. Immigrant inventors may also disproportionately increase the general innovative human capital of their collaborators, allowing them to be more productive even on future teams that the immigrant

New York Stock Exchange.

inventor is not himself a part of. In what immediately follows, we take a reduced-form approach at the individual level and evaluate the extent to which immigrants impact the productivity of their inventor co-authors. Subsequently, in Section 6 we use a structural approach and a team-based analysis to decompose these spillovers into contributions to team-specific capital and contributions to general innovative human capital.

Measuring any given individual's contribution to the productivity of collaborators 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 immigrant and US-born inventors, that is, whether immigrant or US-born inventors 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.[32] 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 US-born inventor 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 productivity spillover effects of immigrant and US-born inventors on their inventor co-authors.

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).[33] 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

[32] 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).

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

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.

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 fifty 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, 3.72 adjusted citations, 0.47 top patents, $65 million of economic value, and a team size of 3.32 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 369,509 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 ten percent are female. Placebo co-authors are, on average, 46.5 years old, with 20 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 6.95 cumulative patents, 1.30 top patents, and received 10.07 total adjusted citations.[34] 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 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.[35] 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:

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

The effects of interest are denoted β_k^{real}, where k denotes time relative to death. These estimates reflect the causal effect of co-inventor death on the outcome of interest k years around death. Note that the joint dynamics around death for both real and placebo survivors is captured by β_{km}^{all}. We also include individual fixed effects (α_i), absorbing individual time-invariant characteristics.[36]

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

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

[34] 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.

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

[36] 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.

Note that this model once again includes event-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 US-born inventors. Finally, we estimate separately the effect of immigrants pre-mature deaths on immigrant co-authors and US-born co-authors, and repeat the same empirical exercise for US-born inventors' 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 collaborator's death, suggesting that the causal effect of an immigrant inventor death on his or her team is larger than that of a US-born 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 immigrant or US-born. The coefficient β^{real} equals to -0.087 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.[37] Relative to this expected mean (reported in Table 4), the treatment effect implies that a deceased inventor lead collaborators to produce 10.3% 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 -0.182 and is again highly statistically significant, implying a 16.3% decline in patenting. By contrast, in column (3) we focus on the causal effect of pre-mature death of US-born inventors, and find that the magnitude of the decline in productivity of co-inventor, as measured by number of patents, while still statistically significant, is only 9.2%. The dynamic treatment effects for US-born and immigrant inventors 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 -0.154,

[37]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.

which is equivalent to a 14.7% 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 23.0% in adjusted citations filed by co-authors, while the effect is only 13.1% for US-born inventors. 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.[38,39]

5.4 Mechanisms driving spillover differences

Having established differential productivity impacts of US-born 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 US-born inventors. 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 US-born deaths. For example, the average dying immigrant inventor is 0.48 years older than the average dying US-born inventor. If we allow for treatment effect heterogeneity of an inventor death based on their age, then our estimated spillover gap between US-born and immigrant inventors would only capture the differential effects above and beyond the treatment effect heterogeneity driven by 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

[38] These results are robust to using the Callaway and SantAnna (2021) methods that deal with possible issues with staggered difference-in-difference research designs. The issues that this estimator fixes are not a problem for our research design since we include match group-X-event year fixed effects for all of our specifications. To be sure, Table A.4 shows we get very similar results using the Callaway and SantAnna (2021) estimator.

[39] One alternative specification is a Poisson estimation, which has the most economically interpretable units. However, since the estimation must be carried out using MLE, having thousands of fixed effects in the regression make it essentially impossible to estimate. For this reason, we have opted to assess robustness using a $\log(1+x)$ specification in Table A.5 in the Appendix: the death of an immigrant co-inventor has a larger negative impact, consistent with our main specification.

the year prior to death. We measure each of these and then convert them to z-scores so that their units are comparable. To implement this, we augment equation (2) as follows:

$$\begin{aligned} Y_{it} &= \beta^{real}_{imm} X^{imm}_{i} * AfterDeath^{real}_{it} + \beta^{real}_{Z_i} Z_i * AfterDeath^{real}_{it} \\ &\quad + \beta^{real} AfterDeath^{real}_{it} + \sum_{k=-9}^{9} \beta^{all}_{mk} \mathbb{1}_{L^{all}_{it}=k} + \alpha_i + \epsilon_{it}, \end{aligned} \tag{3}$$

where X^{imm}_{i} is a binary indicator denoting whether the dying inventor is an immigrant and Z_i is a vector of observable characteristics of the dying inventor and surviving inventors measured in the year prior to death. These are the additional dimensions of treatment effect heterogeneity described above.

We estimate equation (3) and report the coefficients in Table A.6. For example, we find that a one standard deviation increase in the age of the deceased inventor leads to 0.044 *more* patents by the surviving inventor. This implies that inventors who die younger and are thus at a more productive point of their careers have larger impacts on their surviving coauthors. Table (3) reveals other intuitive forms of treatment effect heterogeneity. Row 3 shows dying inventors who were more productive prior to death have larger (more negative) treatment effects on their surviving collaborators. Row 6 shows that the treatment effects are larger (more negative) when the last collaboration between the dying inventor and co-author was more recent. The table reports the treatment effect heterogeneity over 10 different dimensions, which are estimated simultaneously, as denoted in equation (3).[40]

In Table 5, Panel A reports the baseline treatment effect differential between US-born and immigrant inventors before controlling for these additional dimension of observable treatment effect heterogeneity. Panel B reports the treatment effect differential after controlling for the ten observable heterogeneous treatment effects, (reported in Table A.6). 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 US-born and immigrant inventors remains essentially unchanged for scaled citations and economic value, and in fact increases by a statistically significant amount for number of patents and number of top patents.

To investigate why the immigrant-US-born spillover gap does not narrow we estimate a Gelbach (2016) decomposition of the difference based on the ten dimensions of treatment effect heterogeneity.[41] Positive (negative) percentages indicate that controlling for the treatment effect heterogeneity based on the given observable widens (shrinks) the difference. We first see from Row 3 of Panel

[40] While we run these heterogeneous treatment effects for all four of our outcome measures, the estimates using economic value may be less informative because imputation for the privately-own patents economic value were imputed using a regression that not contain all of these values. This could make treatment effect heterogeneity along these dimension misleading. We will not focus on this outcome in this section of the paper.

[41] The Gelbach (2016) method applies the omitted variable bias formula to quantify how much of a focal coefficient from a regression without a given set of controls can be explained by each of the additional controls, once they are added to the regression. See Gelbach (2016) for the full setup and details.

D of Table 5 that accounting for treatment effect heterogeneity driven by the productivity of the dying inventor does narrow the treatment effect gap between immigrant and US-born inventors for number of patents, scaled citations, and top patents. 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 US-born 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 15.7% for the number of patents, 4.8% for scaled citations, and 22.4% for top patents.

However, we also see from Row 6 that controlling for the time since the dying inventor co-authored a patent with the surviving inventor actually widens the immigrant-native gap across productivity measures. Specifically, it widens the gap by 10.6% for number of patents, 4.2% for scaled citations, and 3.6% for top patents. This is due to the fact that the dying immigrants have collaborated with the surviving co-author less recently than dying US-born inventors and that death involving more recent collaborations have more negative (larger magnitude) treatment effects.

Taken together, Table 5 shows that there is meaningful treatment effect heterogeneity that is consistent with economic intuition. Moreover, immigrants and US-born inventors differ on key observable characteristics, such as own and co-author productivity, that generate significant treatment effect heterogeneity. However, collectively these ten dimension of treatment effect heterogeneity do little to explain why immigrant inventor deaths have greater impact on their collaborators' productivity. This provides strong evidence that there is indeed something special and unique about the immigrant inventor 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. The purpose of this section is two-fold. Using a structural framework, we first study whether the disproportionate spillover effects documented in the previous section arise solely from immigrant contributions to team-specific capital, or whether immigrants increase the innovative human capital of their co-authors, which increases their productivity even on teams the immigrant co-author is not a part of. We then further use this structural framework to quantify the share of total US-based innovation that can be attributed to immigrants.

6.1 Team-Based Innovation Framework

We begin by introducing a structural framework in which innovation is produced by teams, where the productivity of the team depends on a team-specific productivity as well as the human capital of the constituent team members. Let j denote a team of inventors collaborating on at least one patent in the sample. We assume that the innovative output of a team j in year t is given by the following production function:

$$
\begin{aligned}
Y_{jt} &= \left(1-\beta^{death}\mathbb{1}^{death}_{jt}\right) h_{jt} g\left(|J_j|\right)\varepsilon_{jt} && (4)\\
h_{jt} &= \left(\prod_{i\in J_j} h_{it}\right)^{\frac{1}{|J_j|}} && (5)\\
h_{it} &= \left(1+N^{imm}_{it}\right)^{\beta^{imm}}\left(1+N^{nat}_{it}\right)^{\beta^{nat}}. && (6)
\end{aligned}
$$

Here $\mathbb{1}^{death}_{jt}$ is a binary variable equal to one if any of the team members of team j are no longer alive in year t and J_j is the set of team members on team j. Thus, if a team member dies, the output of that team in the future is reduced by a constant fraction β^{death}.[42] The variables N^{imm}_{it} and N^{nat}_{it} denote the number of alive unique prior immigrant and US-born collaborators of team member i in year t, respectively. h_{it} denotes the human capital level of inventor i in year t. Equation (6) shows that an individual's human capital is a function of his collaborator network, where we allow differential returns with respect to US-born and immigrant prior co-authors, as reflected by β^{nat} and β^{imm} respectively. The team's innovative productivity then depends on the geometric average of human capital of all members within the team, denoted by h_{jt}. Finally, we allow the team's innovative productivity to depend on team size through a non-parametric function, $g(|J_j|)$, as well as a team-specific idiosyncratic productivity ε_{jt}. If immigrant team members generate high team-specific capital, that will be reflected in higher levels of ε_{jt}.

As evident from equation (6), the death of an inventor affects team productivity through two different channels. First, there is the direct channel by which the team's the dying inventor lowers his teams' productivity by the scaling factor β^{death}. Second, there is the indirect channel. The death lowers the human capital of the dying inventor's co-authors, according to the parameters β^{imm} and β^{nat}. This in turn lowers the productivity of the teams those surviving collaborators are on, even among teams the dying inventor was never a part of. We next discuss how these key structural parameters can be recovered from the data.

Recovering $\beta^{\mathbf{imm}}$ and $\beta^{\mathbf{nat}}$. We focus on number of patents as our dependent variable of interest. We restrict our sample to teams that survived throughout our analysis period, denoted by $\mathcal{J}_1$, such that $\mathbb{1}^{death}_{jt} = 0$ for all t and that β^{death} drops out from equation (6). Intuitively, these are teams that did not experience a direct death of a team member. This subsample allows us to isolate the indirect human capital channel of team members losing co-authors in their prior collaboration network. To derive an estimation equation, we take a first-order approximation of the production functions with respect to the number of unique living immigrant and US-born collaborators.[43]

[42] Theoretically, the productivity of a team should drop to zero after a team member dies, suggesting $\beta^{death} = 1$. However, in the data we observe multiple instances where patents were granted after the team was destroyed due to death, allowing $\beta^{death} < 1$. This is likely due to patent applications and research projects that were initiated prior to a team member's death, but that had not yet resulted in a patent granted at the time of death.

[43] The reason we use a first-order approximation is to simplify issues of dealing with years when inventors have zero output, which prevents taking logs of the production functions. Alternatively, working with the production

Taking this first-order Taylor expansion around a base value and rearranging yields:

$$\left(\frac{Y_{jt}-\overline{Y}}{\overline{Y}}\right)_{j\in\mathcal{J}_1}=\beta_{imm}\left(\frac{N_{jt}^{imm}-\overline{N}^{imm}}{1+\overline{N}^{imm}}\right)_{j\in\mathcal{J}_1}+\beta_{nat}\left(\frac{N_{jt}^{nat}-\overline{N}^{nat}}{1+\overline{N}^{nat}}\right)_{j\in\mathcal{J}_1}+\tilde{\varepsilon}_{jt}, \tag{7}$$

where $N_{jt}^{g}=\prod_{i\in J_j}(1+N_{it}^{g})^{\frac{1}{|J_j|}}-1$ for $g\in\{imm,nat\}$ and all variables with a bar reflect the centering point of the Taylor expansion. Specifically, we consider the approximation around the average values of the placebo-deceased group across years following the inventor death. Appendix B provides a detailed derivation of equation (7).

The left-hand side of equation (7) represents the percent decline in output due to the the change in the team-average number of alive prior collaborators. Thus, to recover the structural parameters, we need to understand how, on average, an inventor death impacts the output of the teams the inventor's collaborators are a part of, excluding the teams the dying inventor is directly a part of. The results are reported in columns (1) and (2) of Panel A of Table 6. The estimating specification resembles equation (2), except now we replace the individual subscript i with the team subscript j. For example, column (1) suggests that the death of an immigrant coauthor in a team member's prior collaboration network lowers that team's productivity by 17.2%. The right-hand side of equation (7) shows that this productivity decline depends on the percent change in the team's average number of immigrant and US-born coauthors. This relationship highlights that, in order to recover the structural parameters, we also need to know how the exogenous death of a coauthor in the network changes the team's average number of prior alive immigrant and US-born collaborators. We estimate these numbers in columns (3)-(6) of Table 6. For example, columns (3) and (5) suggest that the death of an immigrant coauthor in a team member's network lowers the team's average number of prior alive immigrant and US-born collaborators by 25.6% and 7.6%, respectively.

We finally take these estimates from Panel A of Table 6, plug them into the first-order approximation in equation (7), and solve for the two structural parameters of interest.[44] These estimates are reported in columns (1) and (2) of Table 7.[45] We find that the disproportionate immigrant spillover effects documented in the previous section do not solely arise from contributions to team-specific capital. Relative to US-born inventors, immigrant inventors disproportionately contribute to the innovative human capital of their collaborators, human capital which is transferable to

function directly in levels delivers a model where the error term ε_{jt} is non-separable, making structural estimation challenging. The first-order approximation transparently maps standard OLS estimates to structural parameters of the production function.

[44]The $\tilde{\varepsilon}_{jt}$ drops out of the equation since we are plugging sample means.

[45]Specifically, we plug sample means into the LHS and RHS of equation (7), separately for teams experiencing an immigrant death in the co-author network and a US-born death in the co-author network. This provides us with two equations in two unknowns which allows us to estimate the parameters β^{imm} and β^{nat}. First, we use the estimates from Panel A of Table 7 on the effects of an immigrant (or US-born) death on number of patents to calculate $(Y_{it}-\bar{Y})/\bar{Y}$. Second, we calculate the effects of an immigrant (or US-born) death on the team's geometric average number of prior collaborators to estimate $(N_{jt}^{g}-\bar{N}^{g})/\bar{N}^{g}$ for $g\in\{imm,nat\}$. We then use the average team-level number of unique collaborators before death, $\bar{N}^{g}$, to scale this estimate by $\bar{N}^{g}/(1+\bar{N}^{g})$.

teams the immigrant co-author himself is not a part of. In particular, we find $\beta^{imm} = 0.718$ and $\beta^{nat} = 0.290$.

Recovering $\beta^{\mathbf{death}}$. We next restrict our sample to teams that experience an inventor death at some point in our analysis period, denoted by $\mathcal{J}_2$, such that $\mathbb{I}_{jt}^{death} = 1$ for $t \geq t_j^{death}$. We again take a first-order approximation of the production function, this time with respect to team-average human capital and the death indicator, yielding the equation:

$$\left(\frac{Y_{jt} - \overline{Y}}{\overline{Y}}\right)_{j \in \mathcal{J}_2} = \left(\frac{h_{jt} - \overline{h}}{\overline{h}}\right)_{j \in \mathcal{J}_2} - \beta^{death}, \tag{8}$$

where h_{jt} can now be calculated according to the estimates of β^{imm} and β^{nat} above. We again consider the approximation around the average values of the placebo-deceased group across years following the inventor death. Appendix B provides a detailed derivation of equation (8).

The left-hand side of equation (8) represents the percent decline in output due to the death of an inventor on the team. The right-hand side of equation (8) shows that this productivity decline operates through two channels: the direct channel from experiencing a dying inventor on the team and the indirect channel from lowering the team's human capital. Columns (1) and (2) of Panel B of Table 6 display our estimates of the percent decline in output and team-average human capital respectively. We find that the death of an inventor decreases output by 28.2% and human capital by 11.6%. Substituting these numbers into equation (8), we recover $\beta^{death} = 0.166$, as reported in column (3) of Table 7.

Team-Specific Capital. We have shown that immigrant inventors disproportionately contribute to the transferable human capital of their collaborators, relative to US-born inventors. A natural outstanding question is whether immigrants also contribute disproportionately to team-specific capital. In other words, do teams with immigrants have higher levels of ε_{jt}

To study this question, we use our production function parameter estimates to recover the production residual, $g(|J_j|)\varepsilon_{jt}$. We correlate this with the share of immigrants within the team, controlling for the log of team size, $ln(|J_j|)$, yielding the estimates displayed in Panel B of Table 7. Once we strip out team size effects and the spillover effects on collaborators and compare the average innovation across teams with different shares of immigrants, we see that moving from a team that is zero percent immigrants to one hundred percent is associated with a team-specific productivity increase of 5% (0.0036/0.0743). This is statistically significant at the 1% level. Thus immigrants appear to contribute disproportionately to both team specific capital and the transferable human capital of their collaborators.

6.2 Decomposing Aggregate Innovation

We finally use our model to quantify the contributions of immigrant and US-born inventors to total US innovation, taking into account their indirect spillover effects on the human capital of their collaborators. We focus on the number of patents as our metric. We emphasize that these calculations are an accounting decomposition of the observed innovation we see in the data and do not represent counterfactual analysis of what would have happened under different levels of immigration to the United States.

To quantify immigrant indirect contributions through human capital spillovers, we compute the portion of US-born innovation that can be attributed to the human capital generated by their immigrant collaborators. We perform a similar calculation to understand the importance of US-born human capital spillovers on immigrants. To start, we calculate the contribution share of immigrant and US-born inventors by taking all patents granted between 1990-2016, assigning equal credits among all members within the team, adding them up across time and team separately for the immigrant and US-born, and then finally dividing by the aggregate nationwide output.[46] Specifically, for $g \in \{imm, nat\}$, we calculate

$$\frac{1}{Y^{agg}} \sum_j \sum_t \left[\frac{\sum_{i \in J_j} \mathbb{1}_i^g}{|J_j|} Y_{jt} \right], \tag{9}$$

where $\mathbb{1}_i^g$ takes a value 1 if i belongs to group g and 0 otherwise and $Y^{agg} = \sum_j \sum_t Y_{jt}$ is nationwide aggregate innovative output. In column (1) of Table 8, we find that US-born and immigrant inventors produce 77% and 23% of the total patents respectively.

We then calculate how much aggregate innovation would decline if all cross-group collaboration between immigrant and US-born inventors are removed. Specifically, for each patent, we reduce the team-average human capital by setting immigrant team members to have zero US-born collaborators, holding fixed their number of prior immigrant collaborators. Likewise, we set US-born team members to have zero immigrant collaborators, holding fixed their number of prior US-born collaborators. In particular, for $g \in \{imm, nat\}$ and given the production function estimates in Table 8, we calculate:

$$\frac{1}{Y^{agg}} \sum_j \sum_t \left[\frac{\sum_{i \in J_j} \mathbb{1}_i^g}{|J_j|} \widetilde{Y}_{jt} \right] \tag{10}$$

$$\text{where} \quad \widetilde{Y}_{jt} = \left(1 - \beta^{death} \mathbb{1}_{jt}^{death}\right) \left(\prod_{i \in J_j} \widetilde{h}_{it} \right)^{\frac{1}{|J_j|}} g\left(|J_j|\right) \varepsilon_{jt}$$
$$\widetilde{h}_{it} = \left(1 + \mathbb{1}_i^{imm} N_{it}^{imm}\right)^{\beta^{imm}} \left(1 + \mathbb{1}_i^{nat} N_{it}^{nat}\right)^{\beta^{nat}}.$$

Removing cross-group spillovers reduces US-born innovation to 60% of total innovation, as shown in column (2) of Table 8. This implies 17% (77-60) of total US innovation can be attributed to immigrants collaboration spillovers on their US-born collaborators and, further, that 22% ((77-

[46] We focus on the innovative human capital spillover channel and not the team-specific capital.

60)/77) of US-born inventors total innovation can be indirectly attributed to their immigrant coauthors. Removing US-born contributions to immigrant human capital lowers the immigrant share of total innovative output to 15%, implying that US-born human capital spillovers account for 8% (23-15) of aggregate innovation. This further suggests that 35% ((23-15)/23) of immigrant inventors' total innovation can be indirectly attributed to their US-born co-authors.

In sum, these calculations suggest that immigrants contribute directly to 15% of innovation, and their indirect contributions, through the enhancement of US-born inventors human capital, explain 17% of innovation. Together, immigrants account for 32% of total US-based innovation, despite only making up 16% of the inventor workforce. Finally, cross-group spillovers (US-born on immigrants and immigrant on US-born) account for 25% of total US innovation, highlighting the importance of joint US-born and immigrant collaboration in driving US innovation. We summarize these results in column (3) of Table 8.

7 Conclusion

In this paper, we characterize the contribution of immigrants to the innovative output of the United States since 1990. Using inventor address information provided by the USPTO, we link patent records to the Infutor database. We then develop a 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, immigrant inventors are more productive their US-born counterparts, 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 US-born inventors and while they are more likely to work with other immigrants, this tendency declines over time.

Our study raises a number of important questions for future research. First, it would be interesting to examine the heterogeneity of direct and indirect immigrant contributions by country of origin. One could then look at the network effects between immigrants from the same country versus interactions across countries. Alternative data sources might also allow one to study the contributions of immigrants who arrive in the United States as children or the contributions of second-generation immigrants, as compared to the first-generation adult immigrant sample we study here. Another important question is whether immigrants contribute disproportionately to innovation in countries other than the United States. Finally, since our paper is unable to speak to whether immigrants crowd-out innovation of US-born inventors, our paper is unable to provide evidence on how total innovation would change from immigration reform that changes the number of immigrants in the US. This would be a fruitful direction for future research to pursue.

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, "The rise of american ingenuity: Innovation and inventors of the golden age," Technical Report, National Bureau of Economic Research 2017.

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.

Brinatti, Agostina, Mingyu Chen, Parag Mahajan, Nicolas Morales, and Kevin Shih, "The impact of immigration on firms and workers: Insights from the H-1B lottery," *Available at SSRN*, 2023, *4431106.*

Callaway, Brantly and Pedro HC SantAnna, "Difference-in-differences with multiple time periods," *Journal of econometrics*, 2021, *225* (2), 200–230.

Cristelli, Gabriele and Francesco Lissoni, "Free movement of inventors: open-border policy and innovation in Switzerland," *Available at SSRN 3728867*, 2020.

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*, 2019, *109*, 3365–94.

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.

Freeman, Richard B and Wei Huang, "Collaborating with people like me: Ethnic coauthorship within the United States," *Journal of Labor Economics*, 2015, *33* (S1), S289–S318.

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," Technical Report, Social Security Administration 2009.

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.

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.

Wigger, Christoph, "Who with whom? Untangling the effect of high-skilled immigration on innovation," *Journal of Economic Geography*, 2022, *22* (2), 449–476.

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.

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. The blue bars include all patents. The red bars include solo-author patents only.

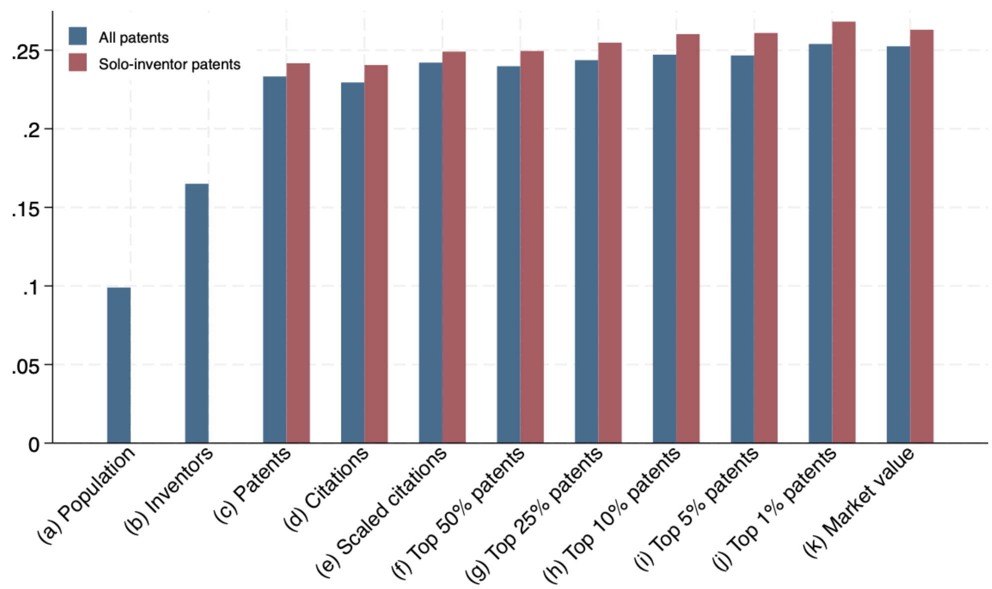
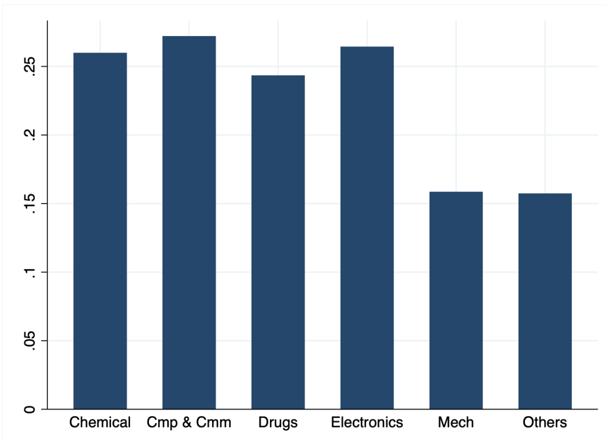


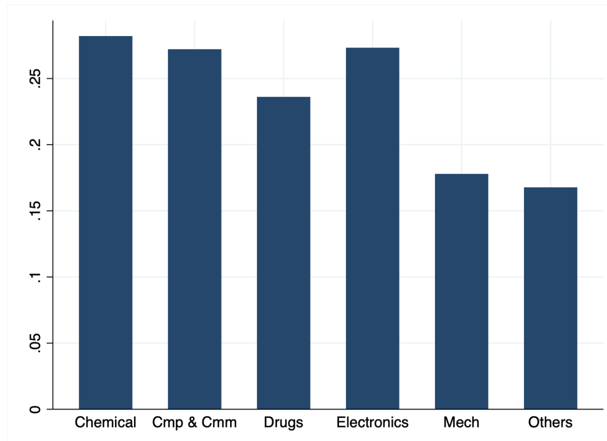
Figure 2: **Share of Immigrant Contribution across Technology Classes**

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.

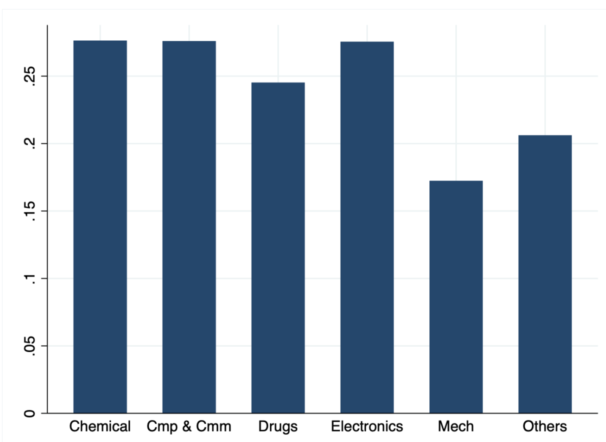
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

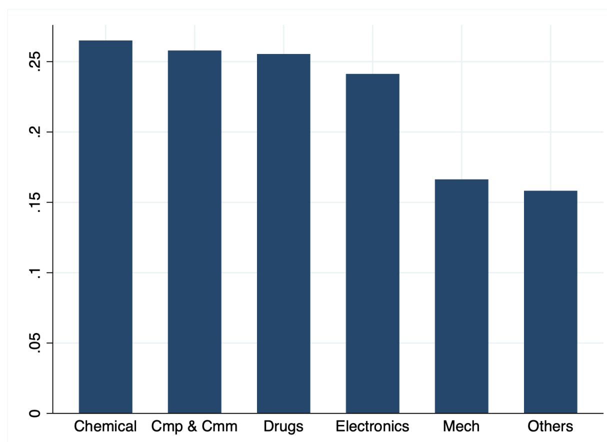
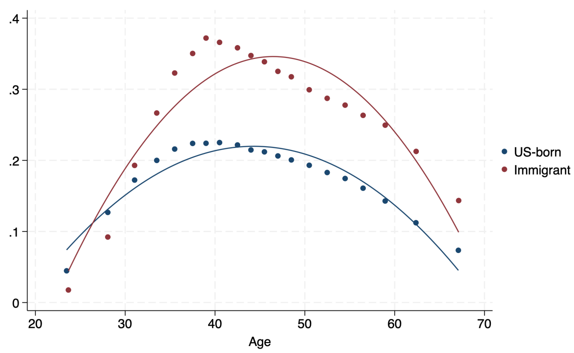


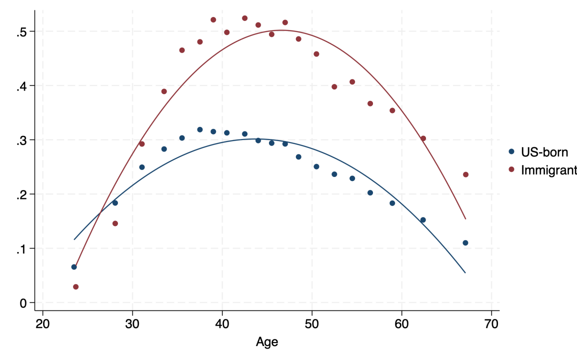
Figure 3: **Productivity over the Life Cycle**

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.

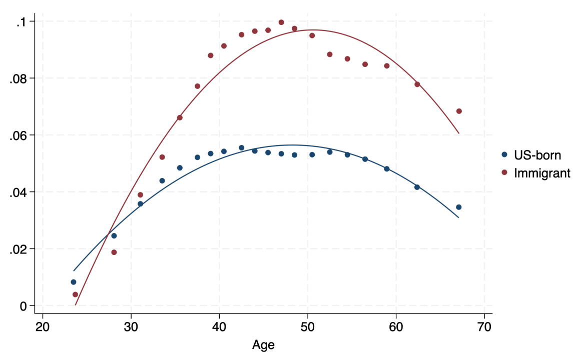
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

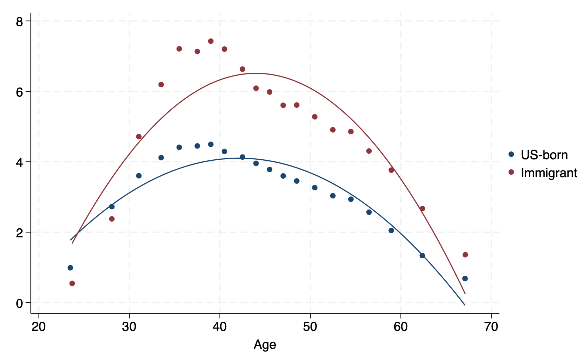
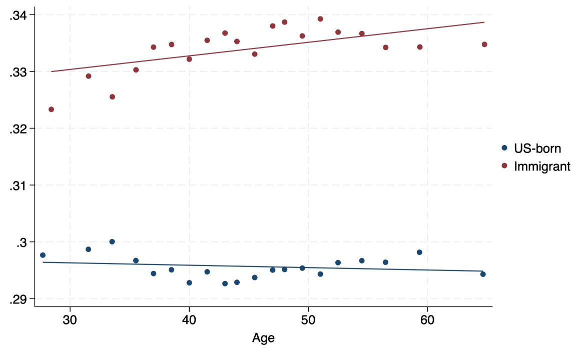


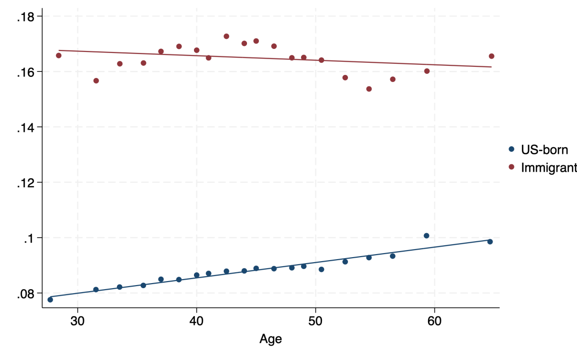
Figure 4: **Global Knowledge Diffusion**

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**

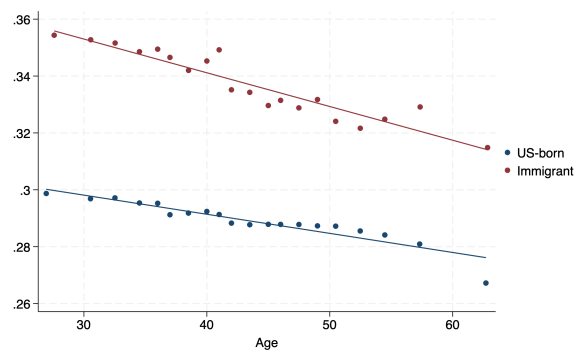
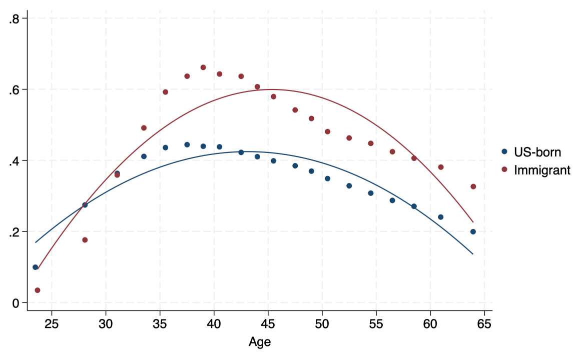


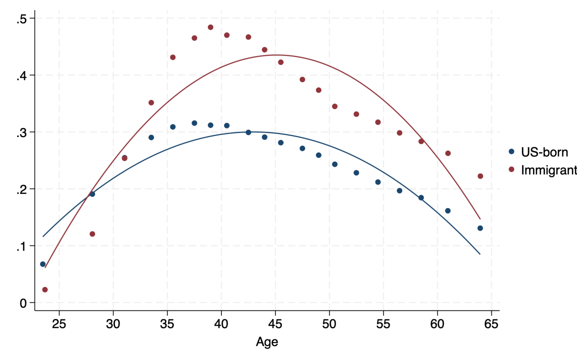
Figure 5: **Assimilation over the Life Cycle**

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.

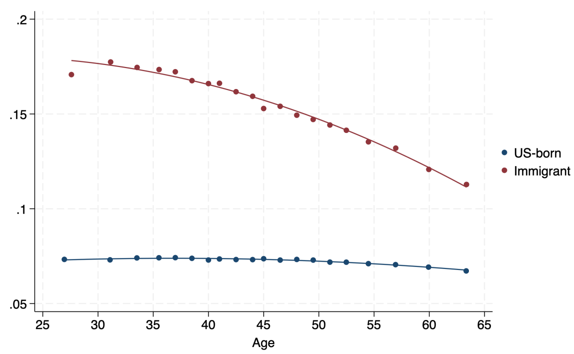
(a) **Number of Unique Co-authors**



(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**

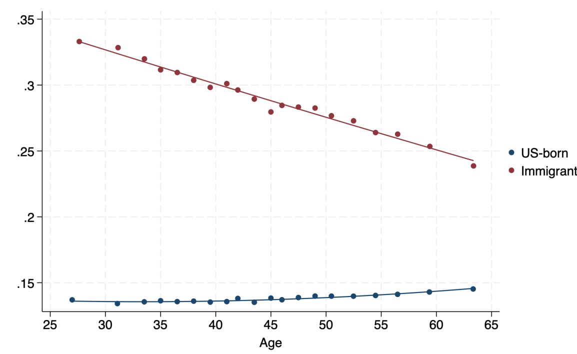
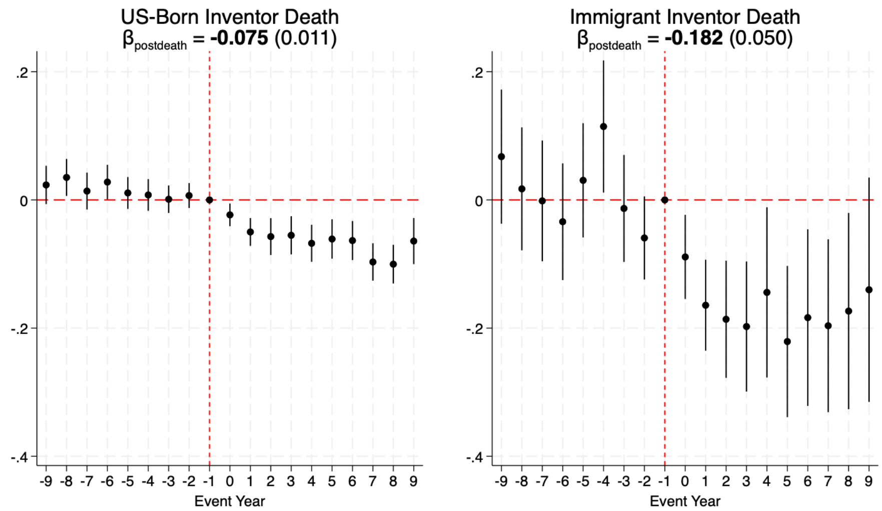


Figure 6: **Comparing Immigrant and US-born Inventor Death**

Notes. Effect of the death of a co-author on inventor productivity for US-born and immigrant inventors, 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**

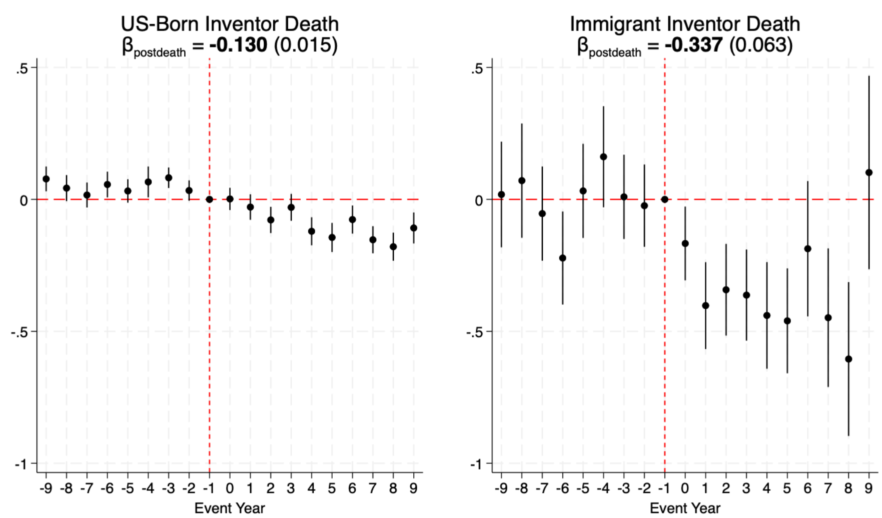
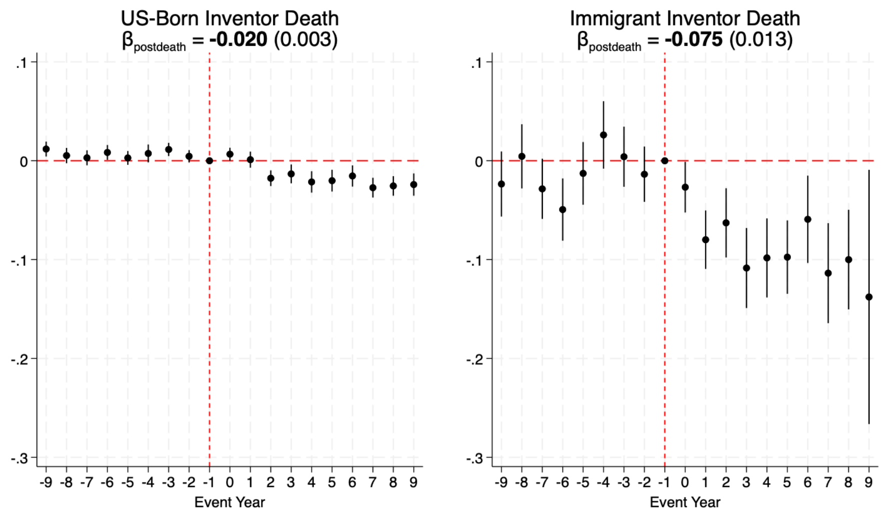


Figure 7: **Comparing Immigrant and US-born Inventor Death**

Notes. Effect of the death of a co-author on inventor productivity for US-born and immigrant inventors, 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**

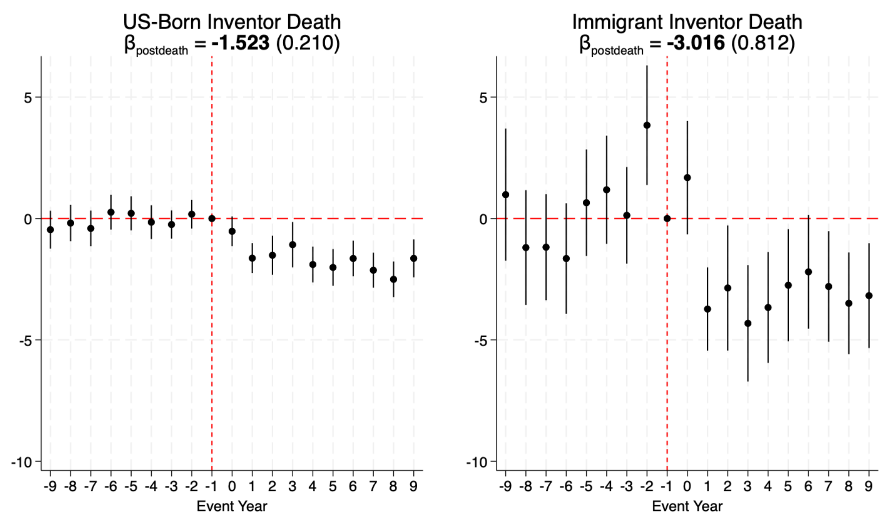


Table 1: **Summary Statistics**

Notes. This table shows summary statistics of the final inventor panel ranging from 1990 to 2016. *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	26.70	187.34	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		652,832
Immigrant	0.17	0.00	1.00		652,832

Table 2: **Productivity Differences between Immigrant and US-born 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. Peer quality is defined as the average number of patents and scaled citations across all coauthors within the team (grouped into ventile bins). 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)	(7)
Immigrant	0.093***	0.092***	0.071***	0.062***	0.071***	0.063***	0.052***
	(0.002)	(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	15,187,669
Year FE	yes	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no	no
County × Tech FE	no	no	no	yes	no	no	no
County × Year FE	no	no	no	no	yes	yes	yes
Tech × Year FE	no	no	no	no	no	yes	yes
Team size × Year FE	no	no	no	no	no	no	yes
Peer quality × Year FE	no	no	no	no	no	no	yes

Panel B: Annual Adjusted Citations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Immigrant	0.149***	0.147***	0.100***	0.086***	0.100***	0.087***	0.072***
	(0.006)	(0.007)	(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	15,187,669
Year FE	yes	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no	no
County × Tech FE	no	no	no	yes	no	no	no
County × Year FE	no	no	no	no	yes	yes	yes
Tech × Year FE	no	no	no	no	no	yes	yes
Team size × Year FE	no	no	no	no	no	no	yes
Peer quality × Year FE	no	no	no	no	no	no	yes

Table 2: (Continued)

Panel C: Annual Number of Top Patents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Immigrant	0.029***	. 0.030***	0.022***	0.020***	0.021***	0.020***	0.017***
	(0.001)	(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	15,187,669
Year FE	yes	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no	no
County × Tech FE	no	no	no	yes	no	no	no
County × Year FE	no	no	no	no	yes	yes	yes
Tech × Year FE	no	no	no	no	no	yes	yes
Team size × Year FE	no	no	no	no	no	no	yes
Peer quality × Year FE	no	no	no	no	no	no	yes

Panel D: Annual Aggregate Economic Value

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Immigrant	1.790***	1.733***	1.213***	0.860***	1.229***	0.945***	0.761***
	(0.054)	(0.054)	(0.055)	(0.055)	(0.055)	(0.056)	(0.053)
Observations	15,714,917	15,714,917	15,714,906	15,192,932	15,709,593	15,187,669	15,187,669
Year FE	yes	yes	yes	yes	yes	no	no
YOB FE	no	yes	yes	yes	yes	yes	yes
County FE	no	no	yes	no	no	no	no
County × Tech FE	no	no	no	yes	no	no	no
County × Year FE	no	no	no	no	yes	yes	yes
Tech × Year FE	no	no	no	no	no	yes	yes
Team size × Year FE	no	no	no	no	no	no	yes
Peer quality × Year FE	no	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 ($\chi^2 = 7.85$). In Panel B, controls include age, immigrant status, and gender for the Infutor matched sample where the characteristics are available ($\chi^2 = 81.68$). 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 ($\chi^2_{deceased} = 8.05$ and $\chi^2_{survived} = 74.95$).

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
Adjusted citations	3.97	1.19	9.38	3.72	1.05	10.74
Top patents	0.50	0	1.40	0.47	0	1.48
Economic Value	76.11	23.11	265.43	64.50	20.79	189.51
Team size	3.37	3	2.20	3.32	3	2.49
Female	0.07	0		0.10	0	
Sample size	3,947			155,711		

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	46.56	48	20.58
Immigrant status	0.15	0		0.20	0	
Co-patents pre-treat	1.91	1	2.26	1.87	1	2.26
Cumulative patents	8.63	3	20.46	6.95	3	16.06
Adjusted citations	12.58	3.24	35.70	10.07	2.39	29.09
Top patents	1.65	0	4.79	1.30	0	3.80
Economic Value	209.40	49.97	585.34	165.03	39.44	539.83
Female	0.10	0		0.11	0	
Sample size	15,471			369,509		

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	68,868	0.08	21,112	0.09	324,931	0.07
Computers	2,843	0.12	108,034	0.12	25,009	0.10	508,505	0.11
Drugs	1,810	0.07	72,390	0.08	16,205	0.07	275,415	0.06
Economic Value	1,957	0.08	79,082	0.09	21,734	0.09	380,562	0.08
Female	1,929	0.08	63,383	0.07	12,953	0.05	224,088	0.05
Sample size	1,936	0.08	63,925	0.07	10,551	0.04	197,966	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	US-born	All	Immigrant	US-born
	(1)	(2)	(3)	(4)	(5)	(6)
Post Death	-0.087***	-0.182***	-0.075***	-0.154***	-0.337***	-0.130***
	(0.011)	(0.050)	(0.011)	(0.015)	(0.063)	(0.015)
Control Post Mean	0.850	1.111	0.815	1.049	1.465	0.993
Percent Change	-10.3%	-16.3%	-9.2%	-14.7%	-23.0%	-13.1%
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
R^2	0.560	0.573	0.557	0.366	0.319	0.377
Number of Deceased Inventors	159,658	8,017	151,641	159,658	8,017	151,641
Observations	6,769,647	502,103	6,267,544	6,769,647	502,103	6,267,544
Dependent Variable:	Top Patents			Economic Value		
	All	Immigrant	US-born	All	Immigrant	US-born
	(7)	(8)	(9)	(10)	(11)	(12)
Post Death	-0.027***	-0.075***	-0.020***	-1.697***	-3.016***	-1.523***
	(0.003)	(0.013)	(0.003)	(0.208)	(0.812)	(0.210)
Control Post Mean	0.201	0.283	0.190	13.307	19.247	12.511
Percent Change	-13.4%	-26.6%	-10.8%	-12.8%	-15.7%	-12.2%
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
R^2	0.373	0.329	0.384	0.454	0.466	0.450
Number of Deceased Inventors	159,658	8,017	151,641	159,658	8,017	151,641
Observations	6,769,647	502,103	6,267,544	6,769,647	502,103	6,267,544

Table 5: **Decomposition of Inventor Death Treatment Effect Differentials**

Notes. This table shows Gelbach decomposition of the treatment effect differentials between immigrant and US-born 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 ten additional characteristics. We standardize each of these characteristics to have mean 0 and standard deviation 1 across all deceased inventors in the sample. Panel C reports the absolute difference between the two panels. Panel D decomposes this difference into percentages explained by these heterogeneity dimension (rows) listed below. Positive (negative) percentages mean controlling for the treatment effect along the given dimension widens (shrinks) the difference. Column (5) of Panel D reports the difference in mean z-scores along these dimensions between deceased immigrants and deceased US-born inventors in the sample. In Panel D, rows 1-3 consider heterogeneity along the deceased inventors' characteristics; rows 4-5 consider heterogeneity along the average surviving coauthors' characteristics; rows 6-9 consider heterogeneity related to the interactions between deceased inventors and their collaborators pre-death; and row 10 considers the number of patents per capita in each commuting zone. In rows 3, 5, and 9, we combine the two sub-dimensions as one by adding and re-standardizing their z-scores. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent variables:	Number of Patents (1)	Scaled Citations (2)	Top Patents (3)	Economic Value (4)	Difference in Mean (5)
Panel A. Differentials *Before* Controlling for Treatment Effect Heterogeneity					
	–0.1065	–0.2073	–0.0548	–1.4923	
	(0.0496)	(0.0620)	(0.0130)	(0.8086)	
Panel B. Differentials *After* Controlling for Treatment Effect Heterogeneity					
	–0.1538	–0.2027	–0.0615	–1.3823	
	(0.0502)	(0.0619)	(0.0130)	(0.8003)	
Panel C. Absolute Difference between Estimates in Panels A and B					
	0.0473	–0.0046	0.0067	–0.1100	
	(0.0113)	(0.0111)	(0.0027)	(0.1753)	
Panel D. Gelbach Decomposition					
Total percentage explained:	44.38%	–2.21%	12.18%	–7.37%	
1. Deceased inventor's age	0.33%	0.31%	0.27%	0.44%	0.0676
2. Deceased inventor's year	–0.31%	–0.07%	–0.18%	0.01%	–0.0502
3. Deceased inventor's cumulative patents and citations	–15.71%	–4.79%	–22.38%	35.06%	0.3196
4. Average surviving coauthors' age	1.44%	0.70%	0.63%	1.54%	–0.0161
5. Average surviving coauthors' cumulative patents and citations	54.03%	–5.03%	29.24%	–29.88%	0.1770
6. Collaboration recency: time to most recent app pre-death	10.56%	4.18%	3.57%	8.57%	0.0697
7. Collaboration network: number of unique coauthors pre-death	–3.45%	–0.43%	4.53%	–23.86%	0.1560
8. Collaboration strength: number of co-patents pre-death	0.60%	0.25%	0.31%	–0.37%	0.0775
9. Collaboration size: average team size on co-patents pre-death	5.19%	2.62%	2.61%	2.70%	–0.1589
10. Knowledge gap: backward citations and overlapping technology classes	–8.31%	0.07%	–6.42%	–1.59%	0.3534

Table 6: **Team-Level Inventor Death**

Notes. This table shows the difference-in-differences OLS estimates of inventor death at the team level that we use to recover the production function parameters. A "team" is defined as a set of unique inventors collaborating on at least one patent in our inventor death sample in Table 3. We classify teams into two subsets. Panel A include teams in which none of member died during the analysis period. However, each member could experience deaths of prior coauthors in the collaboration network outside the team considered. For each outcome in this panel, we consider the effect of immigrant and US-born inventor deaths separately on the number of patents, number of immigrant collaborators, and number of US-born collaborators. Panel B include teams in which one member died at some point during the analysis period. For each outcome in this panel, we consider the overall effect of inventor death, pooling immigrant and US-born deaths, on the number of patents and geometric-average of human capital across all members within the team. Standard errors appear in parentheses and are clustered at the deceased inventor level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Teams That Did Not Experience Inventor Death

Dependent variable:	Number of Patents		Immigrant Collaborators		US-born Collaborators	
Dying inventor:	Immigrant	US-born	Immigrant	US-born	Immigrant	US-born
	(1)	(2)	(3)	(4)	(5)	(6)
Post Death	−0.058***	−0.033***	−1.215***	−0.378***	−0.782***	−1.896***
	(0.019)	(0.005)	(0.090)	(0.022)	(0.152)	(0.049)
Control Post Mean	0.34	0.31	4.75	3.16	10.24	12.08
Percent Change	−17.2%	−10.7%	−25.6%	−11.9%	−7.6%	−15.7%
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Team FE	yes	yes	yes	yes	yes	yes
R^2	0.487	0.434	0.886	0.894	0.888	0.893
Deceased Inventors	5,178	76,755	5,178	76,755	5,178	76,755
Teams	92,138	873,275	92,138	873,275	92,138	873,275
N	1,597,747	15,371,473	1,597,747	15,371,473	1,597,747	15,371,473

Panel B. Teams That Experienced Inventor Death

Dependent variable:	Number of Patents	Team Human Capital
Dying inventor:	Any	Any
	(1)	(2)
Post Death	−0.034***	−0.385***
	(0.003)	(0.013)
Control Post Mean	0.12	3.33
Percent Change	−28.2%	−11.6%
Match Group × Event Year FE	yes	yes
Team FE	yes	yes
R^2	0.288	0.927
Deceased Inventors	103,325	103,325
Teams	181,485	181,485
N	3,202,773	3,202,773

Table 7: **Innovation Production Function Estimates**

Notes. This table shows our innovation production function estimates. In Panel A, columns (1) and (2) display the parameters governing returns to an individual inventor's human capital with respect to the number of unique prior immigrant and US-born collaborators, respectively. Column (3) display the parameter governing the negative productivity shock from experiencing the death of an inventor within the team. Panel B regresses production residuals, averaged across years for each team, on the share of immigrants on that team, controlling for team size. Standard errors appear in parentheses and are clustered at the deceased inventor level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Parameters

Parameter:	β^{imm}	β^{nat}	β^{death}
	(1)	(2)	(3)
	0.718	0.290	0.166
Governing:	human capital	human capital	inventor death
Inventor:	immigrant	US-born	any

Panel B. Residuals

Dependent variable	Team-level Residuals		
Mean dep. var.	0.0743	0.0743	0.0743
	(1)	(2)	(3)
Share immigrant	0.0036***		0.0004
	(0.0005)		(0.0005)
Team size (in logs)	−0.0804***	−0.0803***	
	(0.0003)	(0.0003)	
R^2	0.1319	0.1318	0.0000
N	653,228	653,228	653,228

Table 8: **Decomposing Aggregate Innovation Output**

Notes. This table shows the direct and indirect contribution of US-born and immigrant inventors to total US innovation from 1990-2016. These estimates are based production function parameters reported in Table 7. Innovation is measured in terms of number of patents. Column (1) reports the observed output shares between immigrant and US-born inventors directly in the data. Column (2) calculates output if immigrants only work with immigrants and US-born inventors only work with US-born inventors. Column (3) attributes the indirect effects of US-born and immigrant inventors 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 US-born inventors produce by working with immigrants.

	(1)	(2)	(3)
US-born Output	0.77	0.60	0.68
Immigrant Output	0.23	0.15	0.32
Total Output	1.00	0.75	1.00
Direct Output Attribution:	YES	YES	NO
Indirect Output Attribution:	NO	NO	YES
US-born collaborate with:	Both	US-born	Both
Immigrants collaborate with:	Both	Immigrants	Both

Appendix

A Matching Algorithm of Patent Data with Infutor

In this section, we discuss how we construct our inventor identifiers in the USPTO patent dataset. The idea is to use individuals' address histories in Infutor to verify the address at which each inventor lived during the patent application and track inventors who moved. Precisely, in the patent data, we define "identifiers" (ID) as a unique combination of the inventor's first name, last name, city name, and the state in which he lived. In the raw USPTO data, we begin 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 the Infutor dataset spans 1990 to 2016, we necessarily restrict the patent data to these years only. At the start, there are 6,229,618 patents filed and ultimately 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. Our matching criterion is such that state, city, last name, and the first three letters of first name exactly match between the two datasets: 1,034,288 unique IDs have at least one Infutor match at this stage. In this matched subset, each ID can still map to multiple individuals in Infutor. Thus, we successively apply disambiguation restrictions, where we identify unique matches in a given step and then remove them from the pool before applying the next restriction.

First, we impose that first names match exactly and that middle initials do not conflict between the two datasets (i.e., they agree or at least one or both are 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 names 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 misspellings (e.g., "Stephen" and "Steven"'), while maintaining that middle initials do not conflict. In these steps, we identify 876,438 IDs, each mapping exactly to one individual in Infutor: 680,261, 144,431, and 51,746 IDs from the first, second, and third steps, respectively.

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

In the fourth step, we require that middle initials match weakly and that first names match weakly. The fifth step imposes that first names match strictly, while maintaining middle initials match weakly. In the sixth step, we require that middle initials match strictly and that first names match weakly. The seventh step imposes that both first names and middle initials match strictly. The eighth step considers the cases where both middle names (rather than middle initials) and first names 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. These first eight disambiguation steps together 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 precise 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, completely ignoring first name. In essence, these observations are those with inconsistent first three letters of first name. Then, we condition these potential 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 individuals in Infutor. We summarize all of these data construction steps in Table A.1 below.

A.1 Validation Tests

We begin by comparing the proportion of county-level immigrants based on the entire Infutor dataset and our new classification methodology to the proportion of foreign born individuals at the county level in the 2000 Census.[47] To do so, we first geocode individuals in the Infutor dataset 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

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

well as undocumented immigrants.[48]

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 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.

B Additional Derivations

Derivation of equation (7): $\dfrac{Y_{jt}-\overline{Y}}{\overline{Y}} = \beta_{imm}\left(\dfrac{N_{jt}^{imm}-\overline{N}^{imm}}{1+\overline{N}^{imm}}\right) + \beta_{nat}\left(\dfrac{N_{jt}^{nat}-\overline{N}_j^{nat}}{1+\overline{N}_j^{nat}}\right)$

Since we focus on teams that did not experience a team member death, it must be that $\mathbb{1}_{jt}^{death}=0$ for all j and t. Next, we substitute for members' human capital in the team innovation production function, yielding:

[48] 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 immigrant and US-born individuals based on the age in which they received their social security number.

$$Y_{jt} = \left(\prod_{i \in J_j} \left(1 + N_{it}^{imm}\right)^{\beta^{imm}} \left(1 + N_{it}^{nat}\right)^{\beta^{nat}} \right)^{\frac{1}{|J_j|}} g\left(|J_j|\right) \varepsilon_{jt}$$

$$= \Bigg\{ \underbrace{\left(\prod_{i \in J_j} \left(1 + N_{it}^{imm}\right) \right)^{\frac{1}{|J_j|}}}_{\equiv\left(1+N_{jt}^{imm}\right)} \Bigg\}^{\beta_{imm}} \Bigg\{ \underbrace{\left(\prod_{i \in J_j} \left(1 + N_{it}^{nat}\right) \right)^{\frac{1}{|J_j|}}}_{\equiv\left(1+N_{jt}^{nat}\right)} \Bigg\}^{\beta_{nat}} g\left(|J_j|\right) \varepsilon_{jt}$$

$$= \left(1 + N_{jt}^{imm}\right)^{\beta_{imm}} \left(1 + N_{jt}^{nat}\right)^{\beta_{nat}} g\left(|J_j|\right) \varepsilon_{jt}$$

We take a first-order approximation with respect to N_j^{imm} and N_j^{nat} around a base value $(\overline{N}^{imm}, \overline{N}^{nat}, \overline{Y})$:

$$f(N_{jt}^{imm}, N_{jt}^{nat}) = f(\overline{N}^{imm}, \overline{N}^{nat}) + \left.\frac{\partial Y_{it}}{\partial N_{jt}^{imm}}\right|_{(\overline{N},\overline{Y})} (N_{jt}^{imm} - \overline{N}^{imm}) + \left.\frac{\partial Y_{it}}{\partial N_{jt}^{nat}}\right|_{(\overline{N},\overline{Y})} (N_{jt}^{nat} - \overline{N}^{nat})$$

$$Y_{jt} = \overline{Y} + \left(\frac{\beta_{imm}}{1 + \overline{N}^{imm}} \overline{Y} \right) (N_{jt}^{imm} - \overline{N}^{imm}) + \left(\frac{\beta_{nat}}{1 + \overline{N}^{nat}} \overline{Y} \right) (N_{jt}^{nat} - \overline{N}_j^{nat})$$

Rearranging this yields equation (7) as desired. Note that in practice we calculate the average value across teams in the placebo-deceased group and across post-death years as our base value.

Derivation of equation (8): $\dfrac{Y_{jt} - \overline{Y}}{\overline{Y}} = \dfrac{h_{jt} - \overline{h}}{\overline{h}} - \beta^{death}$

Since we focus on teams that experienced a team member death, it must be the case that (i) for the real-deceased teams we have $\mathbb{1}_{jt}^{death} = 1$ for all j and some $t > t_j^{death}$; and (ii) for the placebo-deceased teams we have $\mathbb{1}_{jt}^{death} = 0$ for all j and t. Given known values of β^{imm} and β^{nat}, h_{jt} is identified. Similarly, we take a first-order approximation of the innovation production function with respect to h_{jt} and $\mathbb{1}_{jt}^{death}$ around a base value $(\overline{h}, \overline{\mathbb{1}}^{death}, \overline{Y})$, which is the average value across teams in the placebo-deceased group and across post-death years (suggesting $\overline{\mathbb{1}}^{death} = 0$).

$$f(h_{jt}, \mathbb{1}_{jt}^{death}) = f(\overline{h}, \overline{\mathbb{1}}^{death}) + \left.\frac{\partial Y_{it}}{\partial h_{jt}}\right|_{(\overline{h},\overline{\mathbb{1}}^{death},\overline{Y})} (h_{jt} - \overline{h}) + \left.\frac{\partial Y_{it}}{\partial \mathbb{1}_{jt}^{death}}\right|_{(\overline{h},\overline{\mathbb{1}}^{death},\overline{Y})} (\mathbb{1}_{jt}^{death} - \overline{\mathbb{1}}^{death})$$

$$Y_{jt} = \overline{Y} + \left(\frac{\overline{Y}}{\overline{h}} \right) \left(h_{jt} - \overline{h}\right) + \left(-\beta^{death} \overline{Y}\right) \mathbb{1}_{jt}^{death}$$

Note that in the post-death period the death effect estimate of $\mathbb{1}_{jt}^{death}$ must equal 1 since it takes a value 1 for the real-deceased teams and a value 0 for the placebo-deceased teams. Recognizing this fact and rearranging this last equation yield equation (8) as desired.

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.

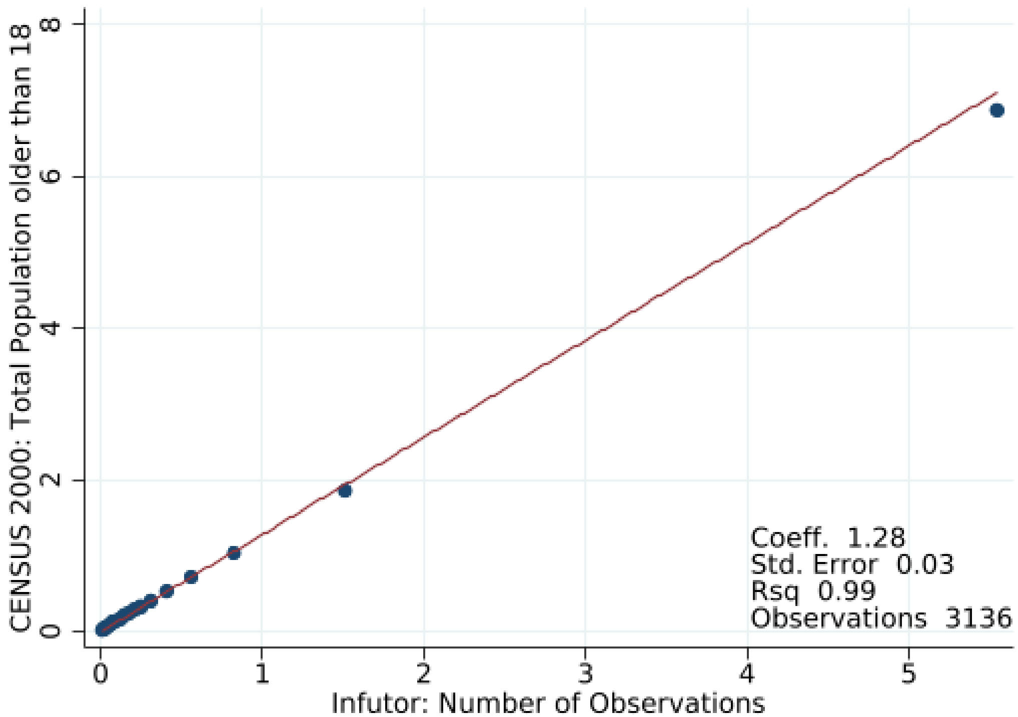


Figure A.2: **Validation of Pre-1950 Assignment Year Imputation**

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.

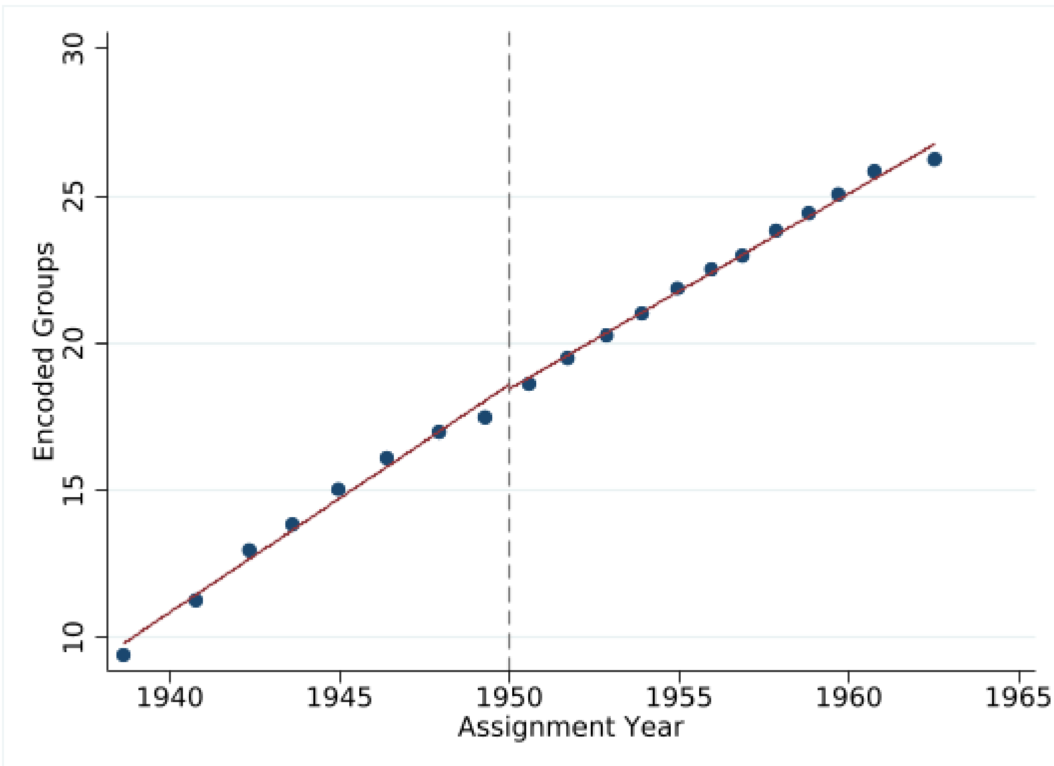


Figure A.3: **SSN Issuance Age Distribution**

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.

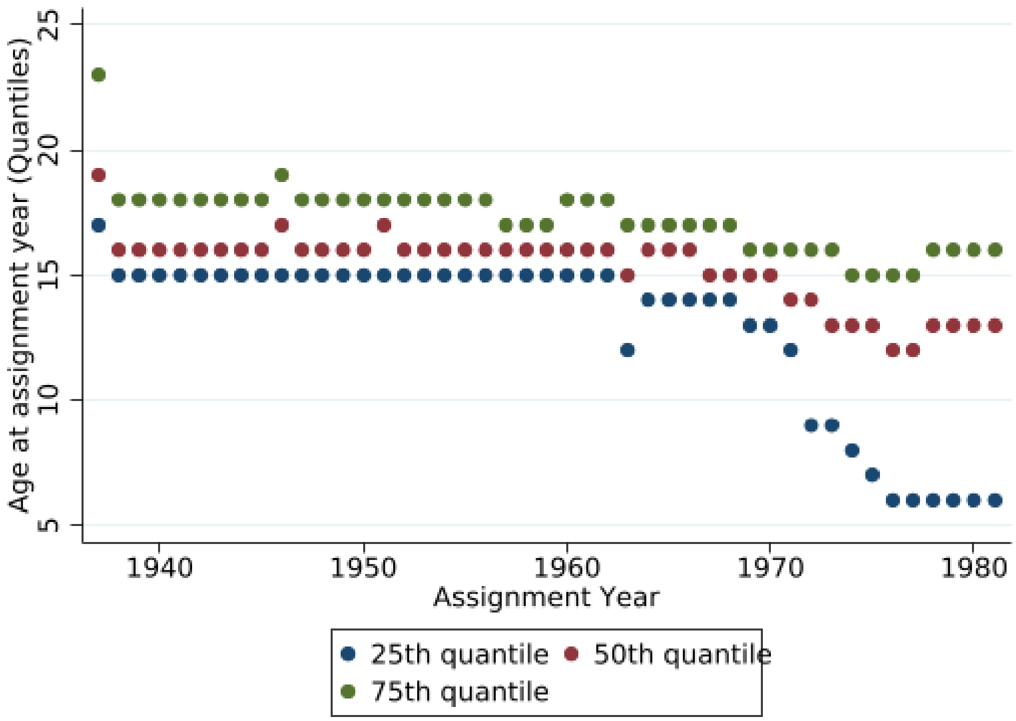


Figure A.4: **Validation with CENSUS 2000**

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.

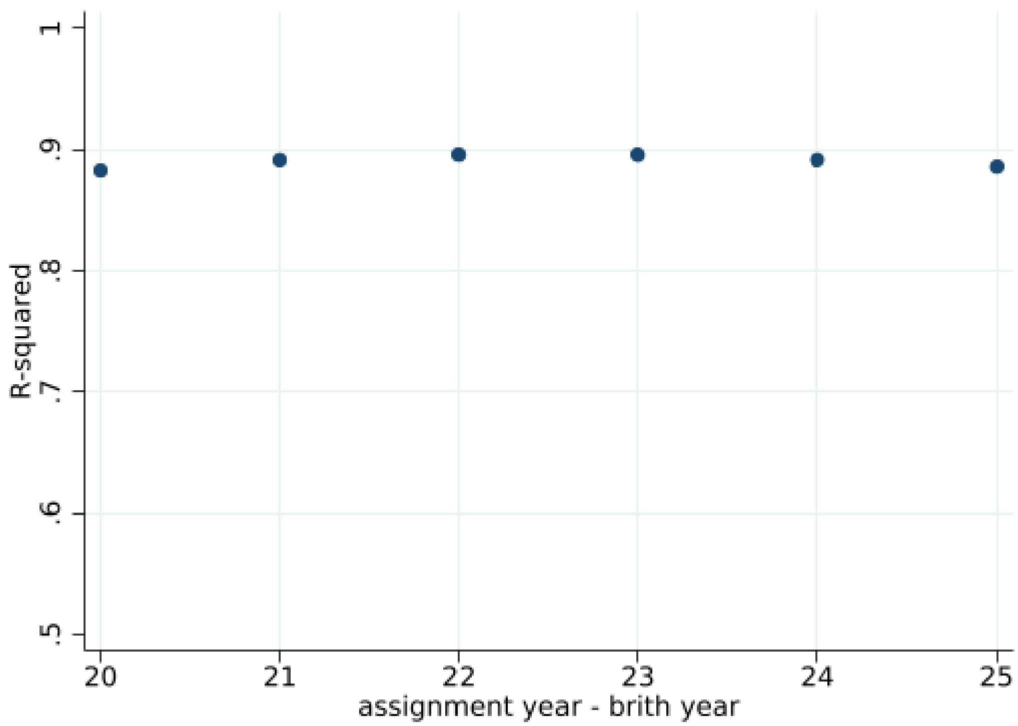
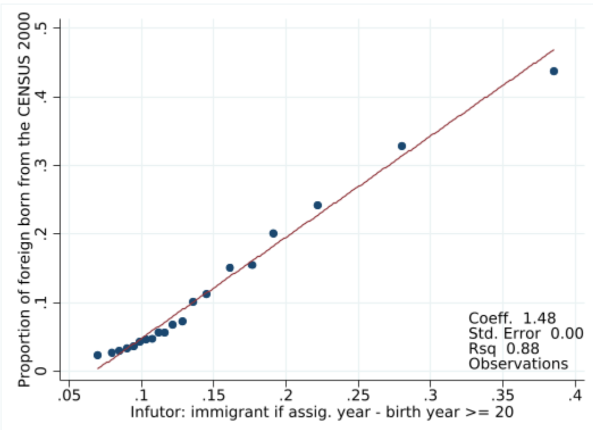


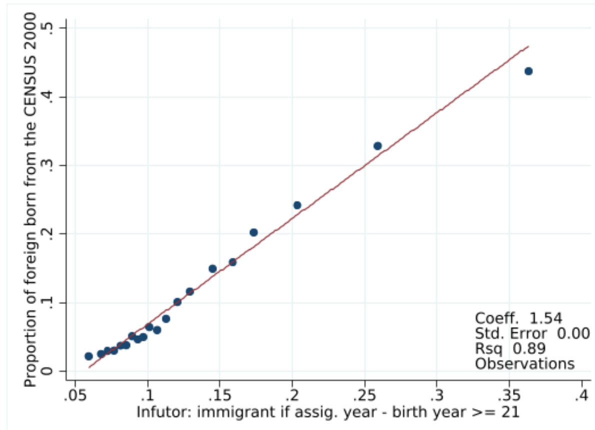
Figure A.5: **Validation with CENSUS 2000 – Binscatters**

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.

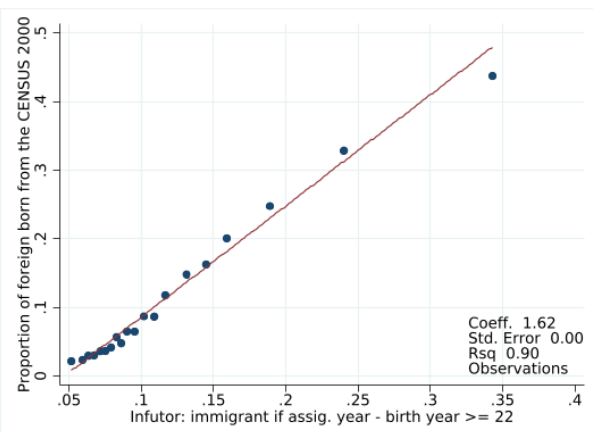
(a) **Immigrant if assig. year − birth year ≥ 20**



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



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



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

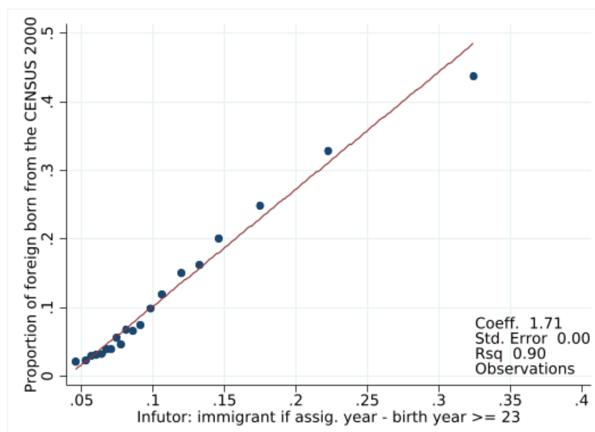


Figure A.6: **Validation with CENSUS 2000**

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.

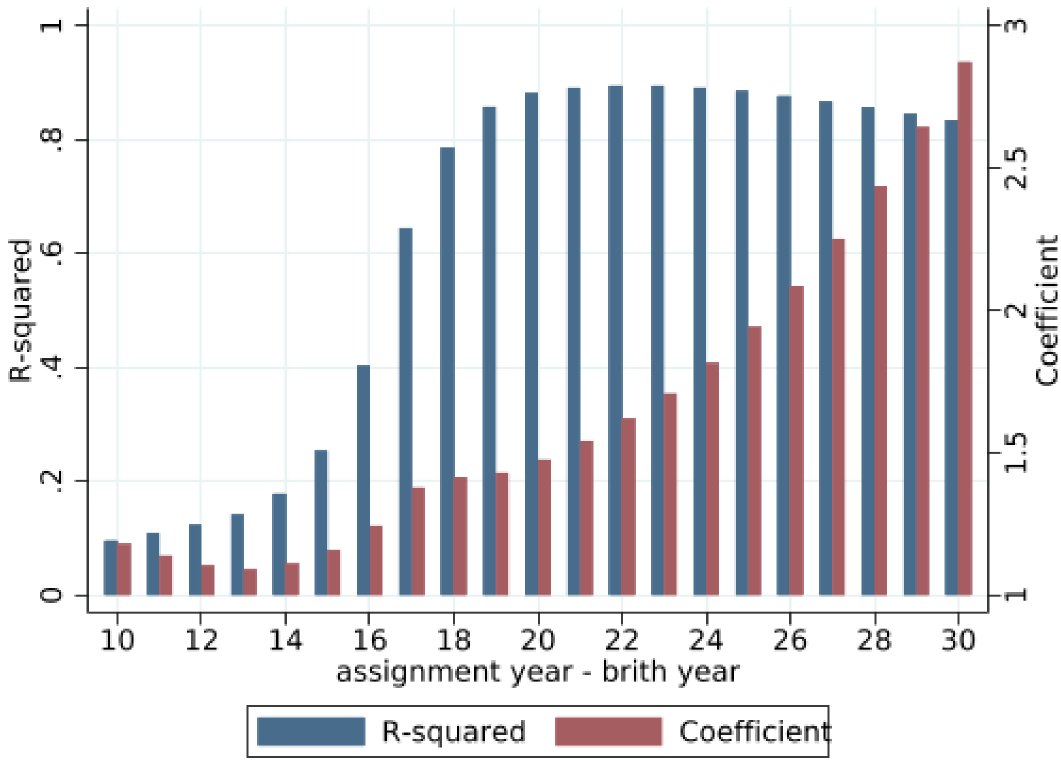
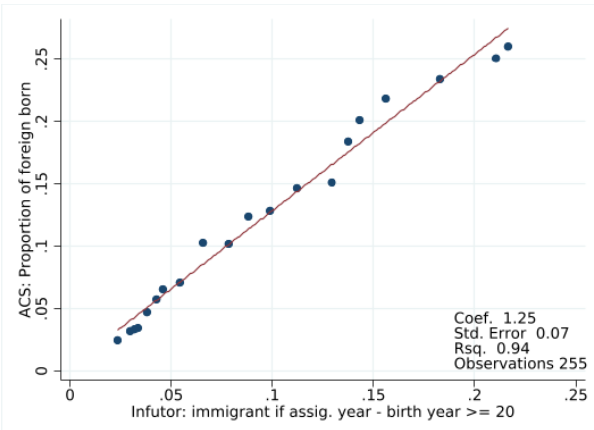


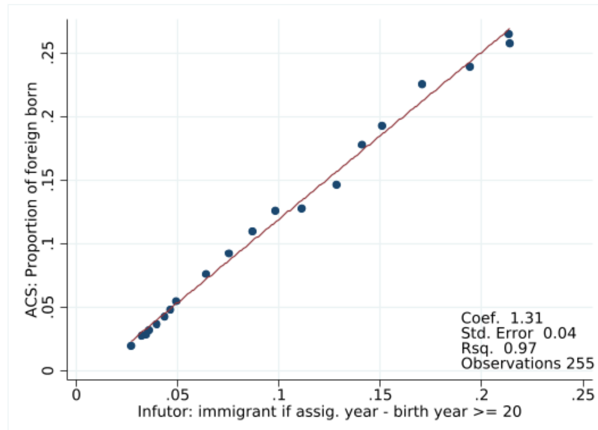
Figure A.7: **Validation against ACS by Selected Age Bins in 2005**

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.

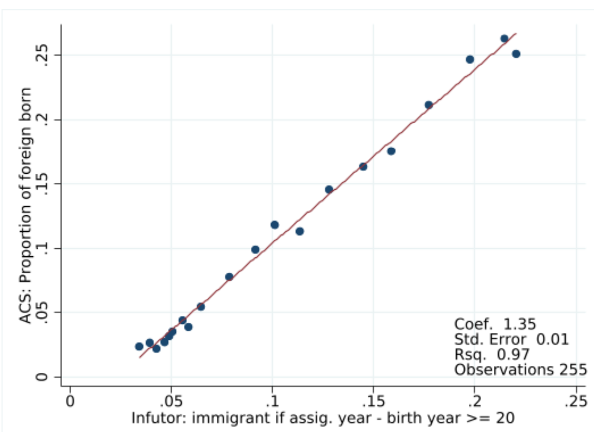
(a) **40-44 years old**



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



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



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

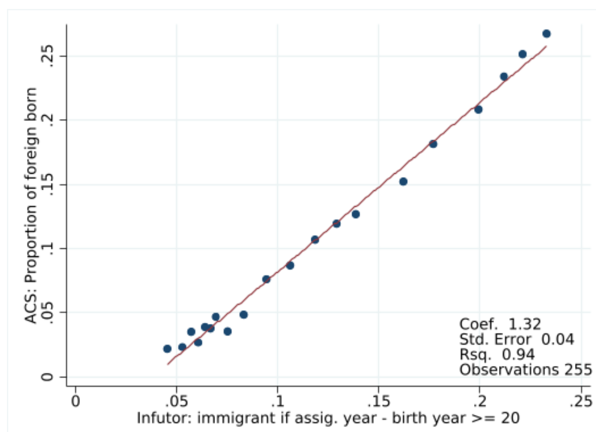


Figure A.8: **Share of Immigrant Contribution, Equal Credits**

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. The blue bars include all patents. The red bars include solo-author patents only.

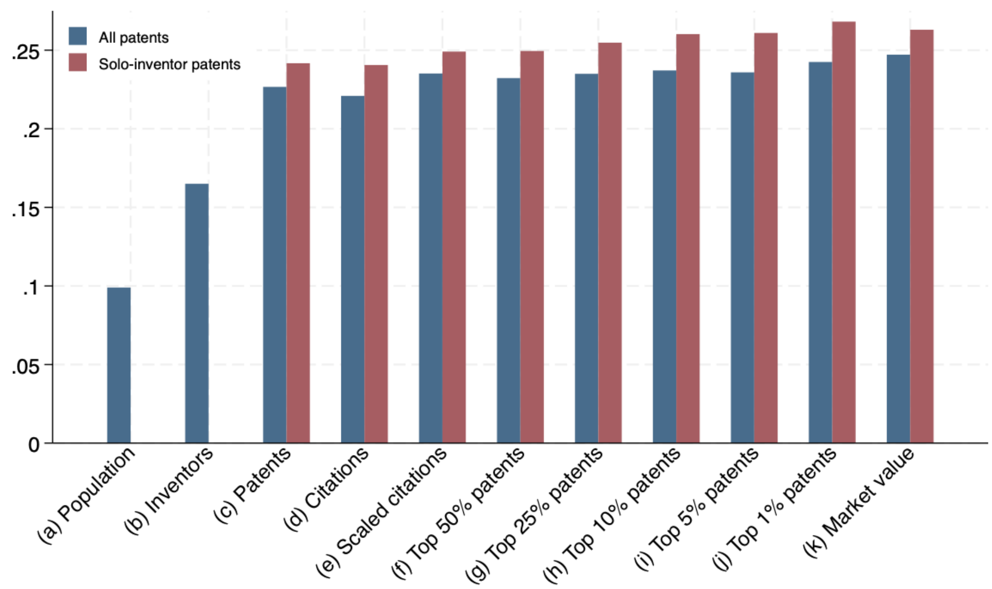


Figure A.9: **Team Size Distribution by Immigration Status**

Notes. Average number of scaled citations by share of immigrant inventors within the team. The plots are displayed separately by team size or the number of inventors on the patent.

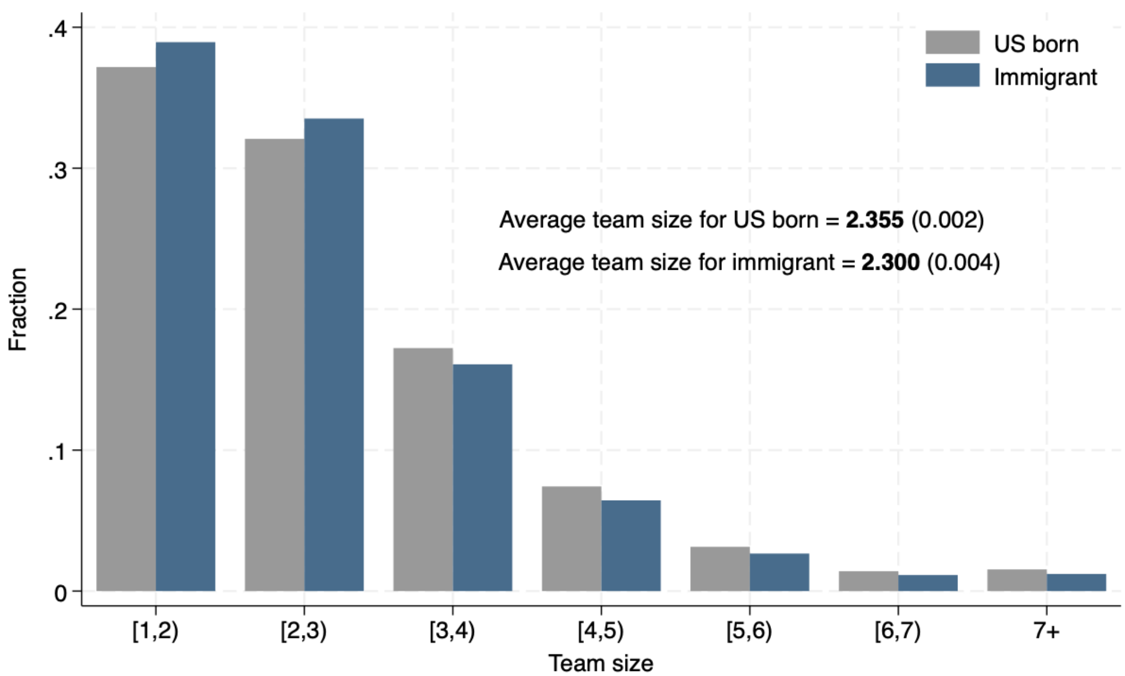


Figure A.10: **Team-level Share of Immigrant Contribution**

Notes. Average number of scaled citations by share of immigrant inventors within the team. The plots are displayed separately by team size or the number of inventors on the patent.

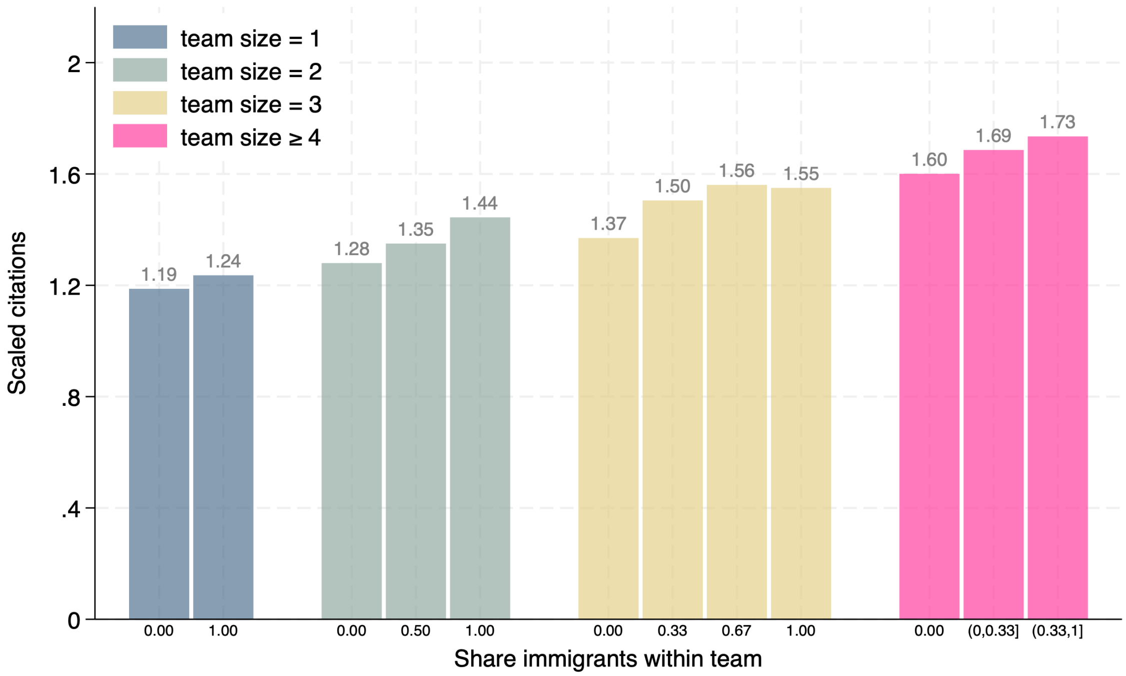
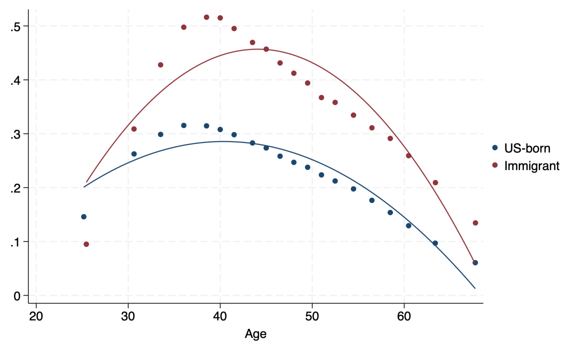


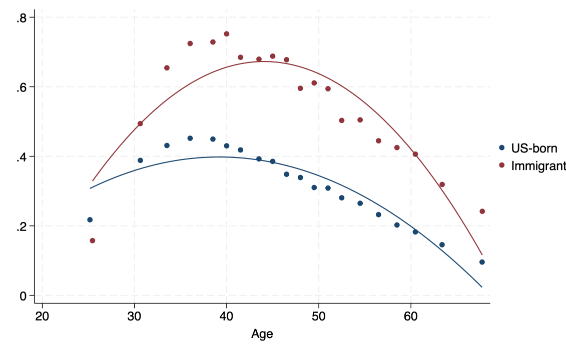
Figure A.11: **Productivity over the Life Cycle - First Patent in 1990s**

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.

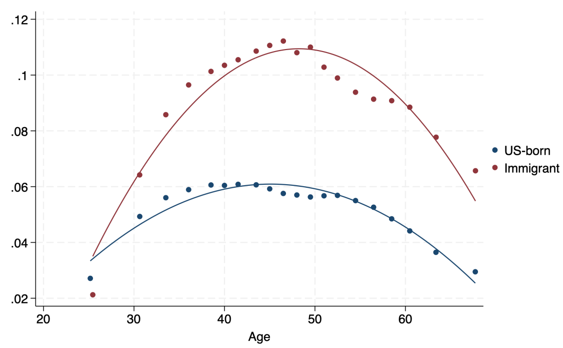
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

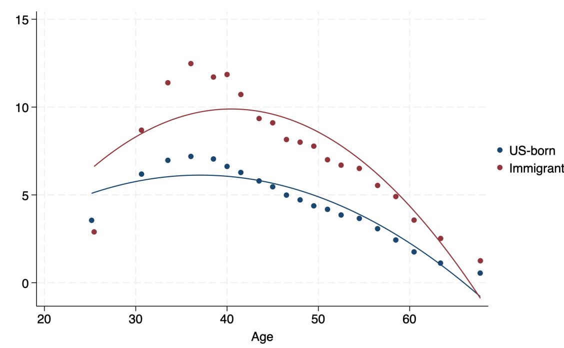
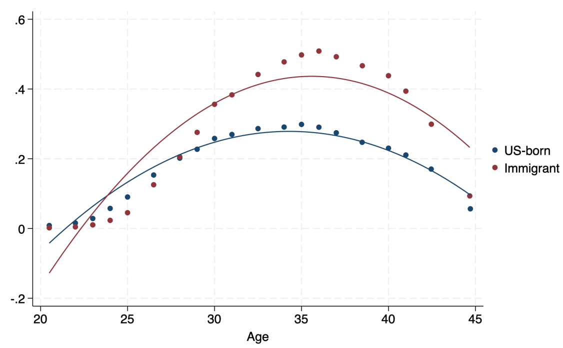


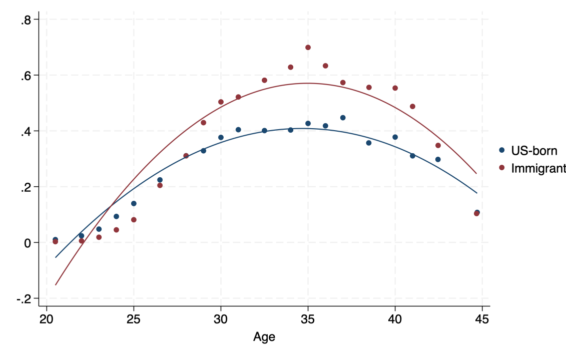
Figure A.12: **Productivity over the Life Cycle - 1970s Year of Birth**

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.

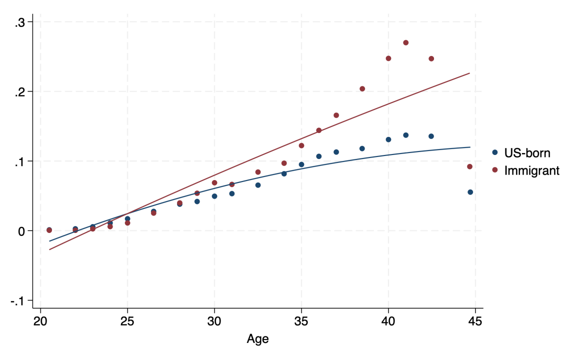
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

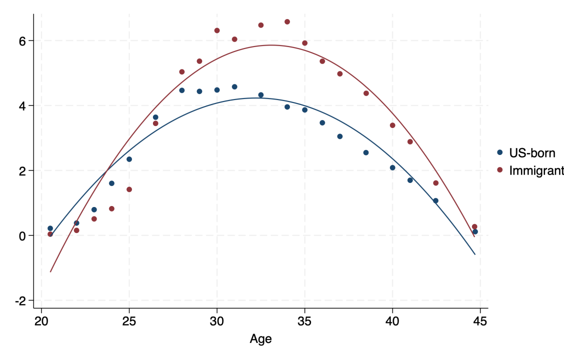
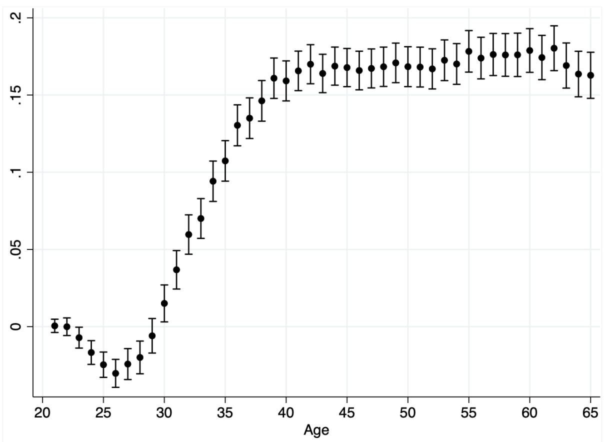


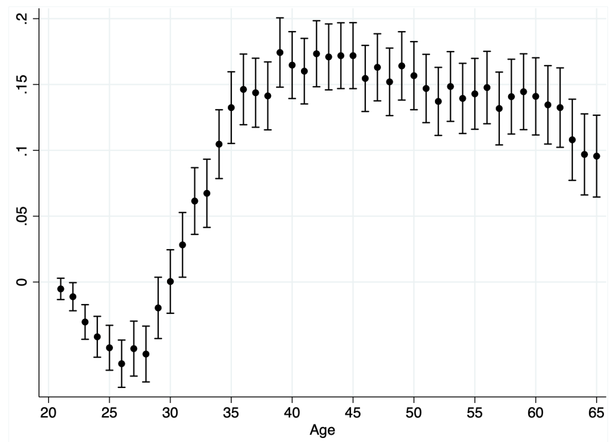
Figure A.13: **Productivity over the Life Cycle - Regressions**

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.

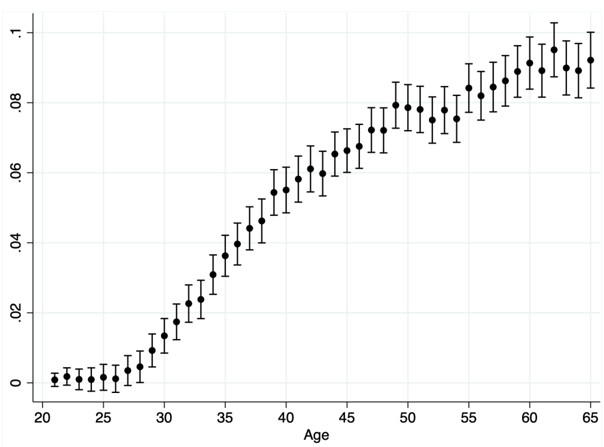
(a) **Number of Patents**



(b) **Total Adjusted Citations**



(c) **Top Patents**



(d) **Market Value**

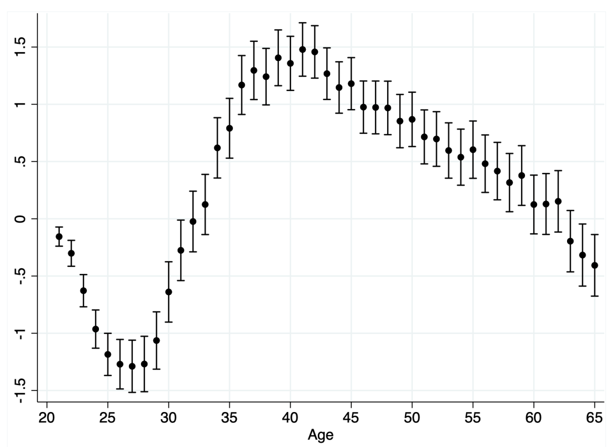


Figure A.14: **Collaboration with US-born and Foreign Inventors**

Notes. The x-axis is the unique number of US-born collaborators an immigrant inventor has over her lifetime. The y-axis is the unique number of foreign collaborators for the same inventor. Since we cannot disambiguate foreign inventors using the Infutor data, we use combinations of their first name, last name, and country at which the patent was applied and granted in the USPTO data to define individual foreign inventors. Our resulting numbers correlate extremely well ($\rho = 0.999$) with those calculated using the inventor identifier from Balsmeier et al. (2015). The regression below controls for year-of-birth fixed effects.

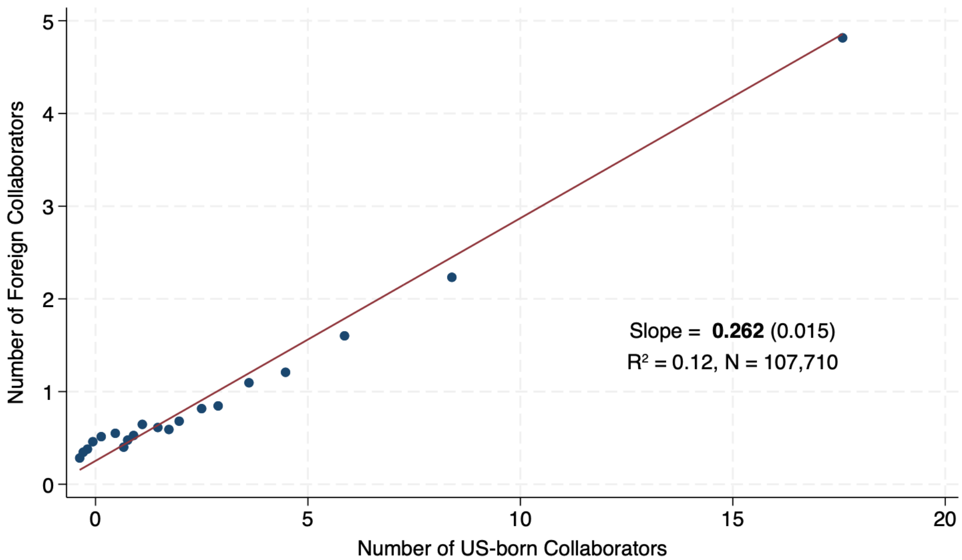


Table A.1: **Sample Construction**

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.

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

Table A.2: **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.3: **Share of Immigrant Contribution – Different Team Sizes**

Notes. This table shows the share of immigrant contribution across different metrics listed in panels (c)-(k) of Figure 1. The statistics are displayed in row 1 for all patents granted between 1990-2016 and in rows 2-6 broken down by team size or the number of inventors on the patent.

Outcome:	number inventors (1)	number patents (2)	raw citations (3)	scaled citations (4)	top 50% patents (5)	top 25% patents (6)	top 10% patents (7)	top 5% patents (8)	top 1% patents (9)	economic value (10)
All patents (100%)	0.165	0.233	0.229	0.242	0.240	0.243	0.247	0.247	0.254	0.252
Team size 1 (54.6%)	0.181	0.242	0.241	0.249	0.249	0.255	0.260	0.261	0.268	0.263
Team size 2 (26.4%)	0.184	0.229	0.227	0.242	0.235	0.239	0.241	0.239	0.245	0.246
Team size 3 (11.3%)	0.185	0.220	0.216	0.234	0.225	0.228	0.230	0.228	0.235	0.245
Team size 4 (4.5%)	0.188	0.212	0.210	0.229	0.216	0.219	0.219	0.218	0.219	0.241
Team size $\geq$5 (3.2%)	0.179	0.202	0.192	0.201	0.206	0.204	0.203	0.200	0.198	0.233

Table A.4: **Robustness to Staggered Diff-in-Diff Methods**

Notes. We replicate our baseline diff-in-diff estimates for the effect on inventor death on surviving collaborators and then repeat the analysis using the Callaway and SantAnna (2021) method.

Specification:	Baseline ATE Results			Callaway & Sant'Anna		
Inventor death:	All	Native	Immigrant	All	Nat	Imm
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A. Number of Patents			
$\beta_{postdeath}$	-0.087***	-0.075***	-0.182***	-0.137***	-0.122***	-0.299***
	(0.011)	(0.011)	(0.050)	(0.017)	(0.019)	(0.044)
			Panel B. Scaled Citations			
$\beta_{postdeath}$	-0.154***	-0.130***	-0.337***	-0.202***	-0.176***	-0.434***
	(0.015)	(0.015)	(0.063)	(0.027)	(0.027)	(0.096)
			Panel C. Top Patents			
$\beta_{postdeath}$	-0.027***	-0.020***	-0.075***	-0.030***	-0.023***	-0.099***
	(0.003)	(0.003)	(0.013)	(0.005)	(0.005)	(0.015)
			Panel D. Economic Market			
$\beta_{postdeath}$	-1.697***	-1.523***	-3.016***	-2.836***	-2.624***	-4.177***
	(0.208)	(0.210)	(0.812)	(0.372)	(0.393)	(1.139)

Table A.5: **Inventor Death, Log Specification**

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:	Log(1 + Number of Patents)			Log(1 + Adjusted Citations)		
	All	Immigrant	US-born	All	Immigrant	US-born
	(1)	(2)	(3)	(4)	(5)	(6)
Post Death	-0.034***	-0.043***	-0.033***	-0.032***	-0.049***	-0.030***
	(0.003)	(0.011)	(0.003)	(0.003)	(0.011)	(0.003)
Match Group × Event Year FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
R^2	0.541	0.548	0.539	0.475	0.480	0.474
Number of Deceased Inventors	159,658	8,017	151,641	159,658	8,017	151,641
Observations	6,769,647	502,103	6,267,544	6,769,647	502,103	6,267,544

Table A.6: **Inventor Death Treatment Effect Heterogeneity**

Notes. Each heterogeneity dimension has mean 0 and standard deviation 1 across the real-deceased inventors in our death analysis sample. Full details are given in the footnote of Table 5.

Dependent variables:	Number of Patents (1)	Scaled Citations (2)	Top Patents (3)	Economic Value (4)
Post-death × treat	–0.056***	–0.103***	–0.014***	–1.286***
	(0.011)	(0.014)	(0.003)	(0.179)
Post-death × treat × immigrant	–0.154***	–0.203***	–0.061***	–1.382*
	(0.052)	(0.064)	(0.013)	(0.828)
Post-death × treat × z-score:				
1. Deceased inventor's age	0.044***	0.079***	0.018***	0.822***
	(0.013)	(0.019)	(0.003)	(0.271)
2. Deceased inventor's year	–0.091***	–0.041**	–0.026***	0.053
	(0.014)	(0.018)	(0.004)	(0.246)
3. Deceased inventor's cumulative patents and citations	–0.046**	–0.027	–0.034***	1.431***
	(0.018)	(0.028)	(0.005)	(0.366)
4. Average surviving coauthors' age	–0.131***	–0.124***	–0.029***	–1.956***
	(0.014)	(0.020)	(0.004)	(0.241)
5. Average surviving coauthors' cumulative patents and citations	0.356***	–0.064	0.099***	–2.755***
	(0.054)	(0.058)	(0.011)	(0.570)
6. Collaboration recency: time to most recent app pre-death	0.164***	0.126***	0.028***	1.860***
	(0.012)	(0.017)	(0.004)	(0.219)
7. Collaboration network: number of unique coauthors pre-death	–0.019	–0.004	0.013**	–1.801***
	(0.020)	(0.030)	(0.006)	(0.377)
8. Collaboration size: average team size on co-patents pre-death	–0.030**	–0.024	–0.008**	0.259
	(0.013)	(0.019)	(0.004)	(0.307)
9. Knowledge gap: backward citations and overlapping technology classes	–0.057***	–0.056***	–0.015***	–0.416
	(0.012)	(0.017)	(0.003)	(0.256)
10. Commuting zone number of patents per capita	–0.021	0.000	–0.008**	–0.055
	(0.013)	(0.018)	(0.003)	(0.260)