The Effects of Higher Education Subsidies on Innovation: Evidence from the GI Bill*

Ran Abramitzky† Gabriele Cristelli‡

September 2024

Preliminary and incomplete
Please do not cite or distribute

Abstract

We study the effects of the 1944 GI Bill on the probability to become an inventor. The GI Bill provided large subsidies to WWII veterans' college education, covering tuition fees at any US university and providing a monthly stipend. Using a fuzzy regression discontinuity design and novel linked inventor-census data, we find a positive effect of GI Bill eligibility on the probability to become an inventor. This result is driven by inventors patenting for companies or government agencies and by inventors active in the computing and electronics fields. We find positive effects also when focusing on inventors patenting high-impact technologies. We provide evidence on the individual role of college education—through GI Bill subsidies—in increasing access to inventor careers in two steps. First, we find a positive effect of GI Bill eligibility on college attendance and on college completion. Second, we find that GI Bill-eligible individuals who resided far from a higher education institution were less likely to access an inventor career. We explore the heterogeneity of our effects by parental background for both inventor and college outcomes. We find a positive effect of GI Bill eligibility for the children of middle-school and college educated individuals, as well as for the children of parents employed as craftsmen or as clerical and sales workers and as professional or technical workers.

*We thank Lukas Althoff, Daniel Fetter, Jenna Kowalski, Thomas Lemieux, Francesco Lissoni, Michela Giorcelli, Paola Giuliano, Santiago Peréz, Hannah Postel, Carlo Schwarz, Heidi Williams, and Gavin Wright for helpful comments. We thank Anna Carolina Dutra Suraiva for excellent research assistance. All errors are ours.

†Stanford University and NBER; ranabr@stanford.edu

‡Stanford University; gcristel@stanford.edu

1. Introduction

Innovation is the engine of modern economic growth (Romer 1990; Mokyr 2005). At its core lie the individuals who advance the knowledge frontier with their scientific and engineering skills (Romer 2001; Jones 2009). Recent research has shown that scientists and inventors are concentrated in narrow segments of the overall population, disproportionately likely to be male, from the ethnic majority, and from high socio-economic backgrounds (Bell et al. 2019; Akcigit et al. 2017b; Aghion et al. 2023; Iaria et al. 2022; Abramitzky et al. 2024a). Differential access to higher education appears to be an important determinant of these disparities, particularly for underrepresented groups by socio-economic background. For instance, conditional on attaining a college education, the patenting gaps by family income significantly diminish (Bell et al. 2019; Akcigit et al. 2017b). Expanding access to university education may thus be a lever to increase the supply of scientists and inventors (Bloom et al. 2019; Van Reenen 2022). Nevertheless, there is limited evidence on the impact of education policies on innovation (Bryan and Williams 2021), particularly on interventions aimed at removing barriers to college access and their effects on the likelihood of pursuing inventor careers.

In this paper, we study the effects of the 1944 GI Bill and on the supply of inventors. The GI Bill provided unprecedented subsidies to World War II veterans' college education, covering the tuition fees at any US university and paying a monthly stipend, which was larger for veterans with dependents and was raised throughout the program.[1] Historians estimate that more than two million veterans attended college using GI Bill benefits (Olson 1974), making it the largest higher education subsidy program in US history. We ask whether the GI Bill increased veterans' probability to become an inventor via access to college education, and how these effects may have differed by parental background.

[1]When the GI Bill was signed into law on June 22, 1944, the monthly stipend was set at $50 for single veterans and $75 for veterans with dependents. These amounts were increased in December 1945 to $65 and $90, respectively. In April 1948, the stipends were further increased to $75 for single veterans, $105 for veterans with one dependent, and $120 for veterans with two or more dependents. The allowance for tuition and books remained constant at up to $500.

Our empirical strategy uses a fuzzy Regression Discontinuity Design (RDD) that leverages variation across birth cohorts in the probability to serve in WWII and to be eligible for GI Bill college aid. Similar to Fetter (2013), our approach is based on the progressive reduction in military conscription for individuals born from 1928 onwards, due to the reduced need of the military as the war ended. Almost 80 percent of US men born between 1922 and 1926 served in WWII, a proportion progressively decreasing for the 1927 cohort, and dropping to 40 percent and below 10 percent for those born in 1928 and 1930, respectively. Intuitively, we compare individuals finishing high-school around the same time and born in the vicinity of a birth year cutoff: those born before 1928, who were highly likely to serve in WWII and to be eligible for GI Bill education subsidies, and those born in or after 1928, who were much less likely to serve in WWII and being allowed to request GI Bill aid.

To conduct our analysis, we construct a novel dataset linking inventors patenting in the US between 1945 and 1990 to the full count 1940 census. We overcome the challenge of linking inventor records generated during adulthood to census information recorded in childhood by gathering birth information for more than half of inventors in our sample. Both the sample of inventors for whom we found birth information and our inventor-census linked dataset are representative subsamples of the full set of inventors patenting between 1945 and 1990 across several key characteristics.

We find that individuals born in cohorts right before the GI Bill eligibility cutoff had a higher probability to become an inventor compared to those born in cohorts right after the cutoff. All our effects are concentrated among white individuals, excluding a positive influence of the GI Bill on inventor outcomes for black individuals. This is consistent with evidence showing that black individuals, particularly in the South, faced limited access to colleges, thereby limiting their ability to benefit from the GI Bill subsidies (Turner and Bound 2003). We find similar effects either when we focus on all inventors in our dataset or when we consider only those who filed their first patent before turning 40 years old, the majority of whom started their patenting

career close to their undergraduate or graduate training. We find larger effects when we focus on inventors in the top quartile of either total career citations, average citations per patent, or highest citations for a single patent. This implies that the increase in the probability to become an inventor was not due only to inventors patenting lower-impact technologies.

Notably, our effects are driven by inventors patenting for companies or government agencies—and not by those patenting in their own name—as well as by inventors active in the computing and electronics fields. These are important results in our context, as inventors in those groups are typically involved in systematic R&D performed at private sector's or federal laboratories, careers typically accessed through undergraduate or graduate training.

Next, we study the heterogeneity of our effects by parental background. We focus on father's education, occupation, and income in 1940, when our birth cohorts of interest were between 10 and 15 years old. When we focus on parental education, we find that our effects are concentrated among the children of individuals with either only middle school education or a college degree. When we consider occupation categories, we find positive effects only for the children of parents employed as craftsmen or as clerical and sales workers and as professional or technical workers. When we consider father's annual income categories, the uncertainty of our estimates is too large to draw conclusive interpretations. This evidence suggests that, among individuals eligible to the GI Bill college aid, the probability to go to college and to access an inventor career increased mainly for those from the lower- and upper-middle class. While most children from upper classes may have gone to college even without GI Bill benefits, several of those born in families from lower socio-economic status may have not.

Our empirical strategy exploits birth cohort variation in the probability to serve in WWII and to be eligible to the GI Bill benefits. As such, it identifies the joint effect of GI Bill eligibility and of military service. We assess the individual role of college education accessed through the GI Bill subsidies to explain our results on inventor outcomes in two steps.

First, we study the effects of GI Bill eligibility via military service on college outcomes.

We rely primarily on newly-released data from the full count 1950 census and link it to the 1940 census for parental background information. We also test effects using a subsample of the 1970 census, providing more granular data by quarter of birth. Our RDD estimates indicate a statistically and economically significant increase in both the probability to attend college and to graduate from college, with the estimated coefficients accounting for one third and 20 percent of the dependent variable's mean, respectively. Similar to inventor outcomes, we find that the effects on college attendance are concentrated among the children of individuals with either a middle school or a high school degree or who were employed as craftsmen or as clerical and sales workers. We also find an effect for the children of individuals in the fourth and fifth quintiles of the income distribution.

Second, we introduce distance from the closest higher education institution in our RDD specification for inventor outcomes. Based on granular geographic residential data from the 1940 census (Berkes et al. 2023), we calculate the distance in miles from the closest four-year university or junior college. Following studies from the literature on the returns to education (e.g., Kane and Rouse 1993; Card 1995; Toivanen and Väänänen 2016), we use distance from a university or a junior college as a proxy of the ease of access to higher education.[2] If the GI Bill college aid played only a minor role in enabling access to inventive careers, we should expect minor differences in the probability to patent among veterans and across birth cohorts, based on their distance to higher education institutions.

Our results indicate that individuals eligible to GI Bill college subsidies who resided far from a higher education institution were less likely to access an inventor career, relative to those who lived closer. This result implies that college education (very likely accessed through the GI Bill) played a role in the determining whether WWII veterans pursued an inventor career.

We conclude our analysis by delving deeper into the effects of the GI Bill and military

[2]In particular, Card (1995) shows that individuals who grew up near a college have higher levels of education compared to those who lived further away. This effect is more pronounced among groups typically characterized by lower educational attainment, such as those from families with low parental education. Hoxby (1997) shows that 93.2 percent of students attended "in-state" colleges in 1949.

service on inventors' sorting across technology fields. Focusing only on inventors, we find that those born in GI Bill-eligible cohorts were disproportionately likely to patent in narrowly defined computing and electronics fields, such as semiconductors, computer hardware and software, and communication technologies.

Our paper contributes to several strands of literature. First, our results add to the literature on the supply of scientists and inventors (Goolsbee 1998; Romer 2001; Bloom et al. 2019; Van Reenen 2022). Recent studies find scientists and inventors disproportionately likely to be male, from the ethnic majority, and from high socio-economic backgrounds (Bell et al. 2019; Akcigit et al. 2017b; Aghion et al. 2023; Iaria et al. 2022; Abramitzky et al. 2024a). There is little or no evidence on interventions increasing access to scientific and inventive careers, particularly for underrepresented groups. The only exception is Bianchi and Giorcelli (2020), who investigate the effects of a 1961 Italian reform, which extended university access to students enrolled in industry-oriented technical high schools, finding a positive effect on their likelihood of filing a patent. To our knowledge, our study is the first to investigate the effects of large individual subsidies to college education, testing the importance of financial barriers to higher education to become an inventor.[3]

Second, we contribute to the literature on access to college education in the US (e.g., Chetty et al. 2020, 2023; Abramitzky et al. 2024b), on higher education interventions (reviews by Deming and Dynarski 2010 and Page and Scott-Clayton 2016), as well as to the broader literature on returns to education (e.g., Griliches 1977; Card and Krueger 1992; Dale and Krueger 2002; Lochner and Moretti 2004; Abramitzky and Lavy 2014). Few studies consider innovation outcomes, only examining the effects of the presence of higher education institutions and its effects on local innovation or the propensity to patent (Toivanen and Väänänen 2016; Dittmar and Meisenzahl 2020; Andrews 2023). Our paper contributes by providing evidence on the link

[3] A closely related literature takes a global perspective and investigates the effects of barriers to STEM carriers for high-achieving individuals in developing economies (Agarwal and Gaulé 2020; Agarwal et al. 2023; Choudhury et al. 2023).

between higher education subsidies, college training, and patenting at the individual level.

Third, we add to the more specific literature on the World War II GI Bill (Bound and Turner 2002; Turner and Bound 2003; Stanley 2003; Fetter 2013; Althoff and Szerman 2022; Eden 2023; Lennon 2023).[4] Bound and Turner (2002) compare men from birth cohorts exposed to WWII to those too young to have joined the military. They find a positive effect of the GI Bill (and WWII service) on the university outcomes of the former group. Turner and Bound (2003) show that such effect was restricted to white men or to black men born outside Southern US states. Using a within-cohort approach, Stanley (2003) finds a positive effect of WWII veteran status on college attainment in cohorts of men going to war right after high school. The effect is noticeably smaller in older cohorts, where veterans were less likely to take up the GI Bill education benefits. Using survey data, Stanley (2003) finds effects driven mostly by veterans from the middle-/upper middle-class.

Our contribution to this literature is threefold. To begin with, our study is the first to provide evidence on the effects of the GI Bill on individual patenting and innovation. Second, we use novel data from the full count 1950 census to study college outcomes, a substantially larger sample than those used by previous studies, which were based on small census subsamples or on the Current Population Survey (CPS). Third, by linking the full count 1950 census with the 1940 census, we are able to study the effects of the GI Bill college aid by socio-economic status at a more granular level than previous research.

Lastly, we contribute to the literature on the effects of WWII on US innovation. Previous studies have examined how wartime R&D investments influenced postwar the rate and geographic locus of postwar patenting (Gross and Sampat 2023), the emergence of the radar industry (Gross and Roche 2023), as well as the effects of invention secrecy orders on patenting during and after WWII (Gross 2023). We add to this literature by studying the supply of the postwar scientific and engineering workforce, focusing on how WWII military service and access

[4] A related set of studies focuses on the Canadian WWII GI Bill (Lemieux and Card 2001), the Vietnam War GI Bill (Angrist and Chen 2011), and the post-9/11 GI Bill (Barr et al. 2021).

to higher education through the GI Bill influenced its size, composition, and sorting across technology fields.

2. Historical Context

2.1. The 1944 GI Bill College Subsidies

On June 22, 1944, President Franklin D. Roosevelt signed into law the Servicemen's Readjustment Act, commonly known as the "GI Bill," a legislation designed to facilitate the readjustment of WWII veterans to civilian life. In addition to low-interest mortgages and loans, the GI Bill offered unprecedented federal support for veterans' college education. The bill covered up to $500 per year in tuition and educational expenses and provided a monthly cash allowance of $50 for single veterans and $75 for married veterans, which increased to $65 and $90 in December 1945, and to $75 and up to $120 (based on the number of dependents) in April 1948. These benefits were available for a minimum of one year to a maximum of four years, with an additional month of benefits for each month of service. Eligible veterans needed to have served at least 90 days of active duty and have been honorably discharged. The majority of veterans qualified for at least three years of benefits, as the average veteran served for 33 months. The benefits were sufficient to cover the tuition fees of most universities in the country. Harvard University's yearly tuition was about $400 in the 1940s (Keller and Keller 2001). Stanford University's tuition was around $429 in 1940 (Mitchell 1946).

Returning veterans had to request GI Bill education benefits to the Veterans Administration and then apply to their chosen college or university, with admissions decisions remaining at the discretion of each institution (Olson 1974). In order to accommodate the increased supply of students, several universities expanded their physical infrastructure—through the addition of temporary and sometimes permanent new facilities—as well as the size of their teaching personnel (Altschuler and Blumin 2009).

The GI Bill saw substantial uptake among veterans, making it one of the largest higher

education support programs in US history. Approximately 2,232,000 veterans attended college supported by the GI Bill (Olson 1973). In the years immediately following World War II, veterans comprised about 70 percent of all male college students in the U.S (Bound and Turner 2002). In 1960, around 18 percent of all college-educated men in the US had their university studies financed by the GI Bill (Stanley 2003).

Historical evidence suggests that a considerable proportion of veterans pursued technical degrees (Geiger 2019). For instance, surveys conducted by the Army and the Navy at the end of the war found that up to 25 percent of veterans planning to attend college aimed to enroll in an engineering program (US Office of War Mobilization and Reconversion 1946). Around 18.5 percent of veterans who enrolled in college by 1952 pursued degrees in engineering, architecture, or physical sciences (US The President's Commission on Veteran's' Pensions 1956). Data from the University of Wisconsin-Madison further shows that veterans were disproportionately more enrolled in engineering degrees than non-veterans (Olson 1974).

The evidence on the effects of the GI Bill college aid by socio-economic status is mixed. Surveys of veterans suggest that the GI Bill may have allowed several individuals outside the elite to access college education. For instance, the Education Testing Service surveyed 10,000 students across sixteen US colleges in the 1946-47 academic year and found that 20 percent of veterans would not have attended college without the GI Bill benefits (Frederiksen and Schrader 1950). More recent studies caution against viewing the GI Bill as an equalizer of higher education opportunity. Using Current Population Survey (CPS) data, Stanley (2003) finds that the GI Bill increased college education attainment only among individuals from the middle and upper middle class

2.2. Selection into the Military During World War II

The US military included both volunteers and draftees during World War II. The Selective Training and Service Act of 1940 established the first peacetime draft, requiring all men aged 21 to 45 to register with local draft boards. In 1942, the lower age limit for the draft was

dropped from 21 to 18. In 1943, volunteering was prohibited, making the draft the sole method of induction for those born from 1925 onwards.

The draft system began with the registration of eligible men at local draft boards, which notified registered individuals to report for duty based on state quotas and "order numbers." Initially, order numbers were assigned through local lotteries. From 1942 onwards, they were assigned based on birth dates, with those born earlier in the year being called for induction first. Inducted men underwent a series of physical and mental examinations. Deferments were granted primarily if draftees were deemed physically or mentally unfit and, in fewer cases, due to employment in critical war production industries or to the presence of dependents. Notably, college students were not exempt from the draft (Flynn 1993).

Deferments meant that WWII veterans were, on average, positively selected relative to their non-drafted peers (Angrist and Krueger 1994). Since volunteering was prohibited from 1943 onwards, it is highly unlikely that individuals selected into the military to benefit from GI Bill once it was introduced in 1944.

2.3. World War II Military Service

Once enlisted, draftees were sent to basic training camps for around 6-12 weeks. Some soldiers received additional specialized training for their assigned roles, which could last from a few weeks to several months. After training, soldiers were deployed overseas or within the US According to Stouffer et al. (1949), as cited in Angrist and Krueger (1994), "out of every four men in the Army one had a combat job and of the remaining three, one had a clerical job, one had a skilled job, and one had a semi-skilled or unskilled job."

Military units were often an heterogeneous mix of draftees by region of origin and personal background. Kennett (1997, p. 60) notes that "[u]nless a man had traveled extensively and done a great many different things in his life, the chances were that the Army presented him with a much broader sampling of humanity that he had ever encountered. While certain units had a lingering regional identity [...] the Army's personnel policies tended to scramble man-

power pretty thoroughly; the thirty-odd draftee divisions formed in 1942 and 1943 were ethnic, cultural, and regional mélanges from the outset."

Thus, military service may have had an ambivalent effect on the lifetime outcomes of returning veterans. On the one hand, the shock of combat for those who served on the front may have had a negative effect. On the other hand, military training, opportunities to build social capital, benefit from mentorship, or simply access to information about education and career prospects, may have had a positive effect.

3. Data

Our analysis requires individual-level data on US residents born in the 1920s and 1930s, with information on their inventor career, education level, WWII military service, birth year, and parental background. In what follows, we describe our sources and data construction pipeline, both for the datasets we use to study inventor outcomes and those to investigate college outcomes.

3.1. Inventive Outcomes

Our analysis of inventive outcomes is based on a dataset linking inventors patenting at the United States Patent and Trademark Office (USPTO) between 1945 and 1990 to the full count 1940 census.[5] A key challenge we face is the absence of birth year and location information in inventor records, which typically provide only inventors' full name and their city of residence as linking features. This complicates linking inventor records generated in adulthood with census information recorded in childhood, which are likely to report different residential locations. We overcome this challenge by retrieving birth information for more than half of inventors in our sample, enabling us to link inventors active during the second part of the twentieth century to their younger selves in the 1940 census, combining information on inventors' careers and productivity with data on their childhood and parental background. This is the main data

[5]Despite their well-known limitations (Griliches 1990), patent records are a key measure of innovation in advanced economies since at least the early twentieth century (Mansfield 1986; Cohen et al. 2000; Moser 2016), providing extensive information on scientists' and engineers' identity and their inventive activity.

contribution of our paper. Previous economics research only linked inventors active close to census years, based on name matches and corresponding residential information (Akcigit et al. 2017a,b; Sarada et al. 2019). This method limited the sample to inventors patenting until the early 1940s, born primarily until the first decade of the 20th century. By using birth information, we extend the linkage to inventors born in the 1910s, 20s, and 30s, who represent a substantially larger portion of those patenting in the post-WWII period.[6]

The following paragraphs describe the main steps in our data construction process, starting with the inventor disambiguation, obtaining inventors' birth information, producing inventor-census links, and adding WWII veteran information.

Inventor disambiguation. The first step in our procedure involves attributing unique identifiers to the same inventor across multiple patents, a practice commonly referred to as disambiguation or entity resolution (Torvik et al. 2005; Trajtenberg et al. 2006; Li et al. 2014; Pezzoni et al. 2014). Inventors from "modern" patents (i.e., granted since 1975) are assigned unique identifiers by the USPTO using an algorithm based on supervised classification and hierarchical agglomerative clustering (Monath et al. 2021). We adopt the same machine learning algorithm and disambiguate also inventors listed on "historical" patents (i.e., granted before 1975; Andrews 2021), assigning a unique identifier to all inventors filing a patent at the USPTO and residing in the US between 1920 and 2015.[7]

We describe our algorithm in detail in Appendix B.2. In brief, we first train a logistic classifier using a set of 2 million disambiguated inventor-patent instances from modern patents and fine-tune it on a development set of additional 100,000 disambiguated records. Our features are based on the inventors' name, residential location, co-inventors' and assignees' identity, and technology classes. Second, we use the classifier's predictions to construct a set of distance ma-

[6]Using our birth information dataset, we find that among inventors who filed patents between 1945 and 1990, 15 percent were born before 1910, while 65 percent were born between 1910 and 1939.

[7]For patents granted before 1975, we combine data from the Patstat database (inventor and assignee name), the PatentCity database (inventor and assignee location; Bergeaud and Verluise 2024), the USPTO historical masterfile (patents' grant date and technology class; Marco et al. 2015), and Google Patents (patents' filing date). For patents granted since 1975, we obtain the same information from the USPTO PatentsView database (USPTO 2022).

trices, each including all inventors sharing the same last name and first name two initials. Third, we apply hierarchical agglomerative clustering to each matrix, assigning the same identifier to inventors reporting distance scores within the optimal threshold that maximizes precision and recall, as determined at the fine-tuning stage.[8]

We assess our algorithm's performance in three steps. First, the optimal threshold determined at the fine-tuning stage achieves precision and recall of approximately 0.99 and 0.95, respectively. Second, when focusing only on inventors patenting since 1975, our inventor identifiers show between 96 and 99.5 percent similarity with those generated by the USPTO, indicating a virtually identical disambiguation. Third, similar to Akcigit et al. (2022), we search for the top 50 most prolific inventors in our dataset in a crowdsourced list of the most prolific inventors known, maintained on Wikipedia. We find 45 inventors in the list and confirm the identities of the remaining five by consulting their biographical profiles on companies' or universities' websites. In addition, for inventors whose entire careers are covered by both our dataset and Wikipedia, we find minimal differences in their total patent stock.

Our algorithm and similar disambiguation approaches achieve high rates of precision and recall for two reasons. First, patent documents provide several key pieces of information to differentiate homonyms, such as residential information at the city level, the identities of collaborators and assignees, and highly granular technology field classification. Second, few inventors patent more than two or three times in their career (in our sample, 80 percent of inventors file fewer than four patents during their lifetimes). For those who file multiple patents, the algorithm would struggle to disambiguate only inventors who simultaneously patent in different locations, with different collaborators and assignees, and in technology fields far from each other.

We argue that any of these mistakes in our disambiguation procedure would most likely

[8]This disambiguation strategy is similar to the method used by Akcigit et al. (2022), who disambiguated inventors from historical patents using a training set of inventors from modern patents disambiguated by Li et al. (2014).

introduce a conservative bias in our estimates by reducing the number of inventors linked, thereby diminishing the statistical power of our estimates. This could occur either by wrongly clustering an inventor with another, missing one of them from our inventor birth information search sample, or by incorrectly splitting records pertaining to the same individual, which would be dropped at the linkage stage because generating non-unique inventor-census pairs.

Inventor birth information search. The second part of our data construction process involves obtaining information on inventors' birth year and location. We start by selecting all inventors in our dataset which patented at least once between 1945 and 1990. We select this patenting time period to capture the majority of the inventive careers of individuals born between the late 1910s and the early 1930s, the birth cohorts of interest for our analysis.

For each inventor in the dataset, we search for their first, middle, and last name, along with their city of residence on *ancestry.com*. We collect only records reporting the same residence as the inventor, at either the state, county, or city level, depending on the granularity of the geographic information available. Informed by studies on inventors' lifecycle (Jones 2009; Bernstein et al. 2022; Kaltenberg et al. 2023), we keep only records referring to individuals who were between 25 and 65 years old during their patenting career. Additionally, we discard any record reporting a death year before the inventor's last patent. Lastly, we retain birth information only for inventors with a unique candidate birth information profile.

Using this procedure, we identify birth information for around 56 percent of inventors in our sample (549,768 out of a total of 985,636 inventors). Birth year information is consistently available. Birth location data is available for 60 to 70 percent of inventors born until 1925, a proportion gradually decreasing for younger birth cohorts (roughly 50 to 35 percent for inventors born between 1930 and 1935). Below, we discuss and address the implications of this uneven birth location coverage for our inventor-census linkage procedure.

Two pieces of evidence indicate that our inventor birth information data is of high quality. First, we manually check a random sample of 300 inventors, finding correct birth information

for roughly 95 percent of those records. In particular, we look for information on the inventors' residence, occupation, industry, and employer, making sure it corresponds to the information found on their patent documents. We also disregard any record which may be associated with homonyms sharing the same birth year and location, as well as the same state of residence. Second, we find that the subset of inventors for whom we found birth information is a representative subsample of the full set of inventors patenting between 1945 and 1990 across several key characteristics. Appendix B.3 reports further details on the inventor birth information search and the tests we carried out to evaluate the data quality.

Inventor-census linkage. The third step in our data construction process involves linking the inventors in our sample to the full count 1940 census. We adopt the class of ABE algorithms (Abramitzky et al. 2012, 2014, 2021a), building on their linkage routine to leverage the information provided by both the census and inventor data.

We focus on inventors born between 1915 and 1935 (235,708 inventors) and proceed as follows. First, we create a set of candidate links by matching inventor and census records on first and last name. Second, we require middle name initials to match, if they are available for both records. Third, we keep only records within a $\pm$ 1 birth year band. Fourth, we require the birth state to match. Fifth, we retain only unique inventor-census records pairs (that is, we exclude inventor records with multiple census candidates or multiple inventors linked to the same census candidate). Sixth, to address the progressively lower availability of birth location information for inventors in younger birth cohorts, we add to the sample unique inventor-census pairs with matching middle name initials (or absent middle name for both) which would have been discarded at the previous step due to the lack of birth location information for the inventor. We obtain a sample of 79,019 inventors linked to the census, around 34 percent of our starting inventor set. We refer to this sample as "linked sample (a)."

To test the robustness of our analysis, we produce two alternative linked samples. For the first ("linked sample (b)"), we follow the same procedure up to step five, keeping only unique

inventors-census pairs sharing the same birth location. This sample includes 43,679 inventors linked to the census (or 19 percent of the starting inventor set). For the second ("linked sample (c)"), we follow the same procedure up to step three and keep only unique inventor-census pairs with matching middle name initials (or absent middle name for both). This sample includes 74,286 inventors (or 32 percent of the starting inventor set).

We assess the representativeness of each linked inventor sample by comparing some of their key characteristics with the starting inventor set we aimed to link. Both linked sample (a) and (c) show minimal differences relative the starting inventor set, displaying similar average birth year and lifetime patents, as well as highly comparable proportions of inventors across first patenting regions and by technology category (Appendix Table B5). Records in the linked sample (b) show two key differences compared to our starting inventor set. First, they are relatively more represented in the mechanical engineering category and among inventors patenting in the Midwest. Second, they report a slightly older year of birth (1923.7 compared to 1924.9 for the full starting sample).

Such birth year difference is evident when comparing the entire birth year distribution of all inventor samples (Appendix Figure B5). The distribution of inventors in linked sample (a) and (c) closely tracks that of our starting inventor set. Linked sample (b) has a more skewed distribution, less representative of inventors born from 1928 onwards. This is explained by the linkage routine generating this linked sample, restricting links to those with available birth location, which are progressively fewer for younger birth cohorts.

WWII veteran information. The last step in our procedure involves adding information on WWII military service to the 1940 census. We use the WWII Army Enlistment Records dataset, which includes around 9 million records of individuals who served in the US Army and Army Air Corps during WWII. This dataset originates from punch cards generated when enlistees joined the Army, providing information such as the enlistees' name, birth year and location, ethnicity, date and place of enlistment, and army grade and branch. We use links

between the WWII Army Enlistment Records dataset and the 1940 census produced by Wikle and Osborne (2023), who linked around one third of enlistees to the census employing the ABE algorithm (Abramitzky et al. 2012, 2014, 2021a).

We classify all individuals linked to the WWII Army Enlistment Records as WWII veterans. It is important to note that this measure identifies only a portion of WWII veterans in our dataset, for three reasons. First, the enlistment records include only Army and Army Air Corps enlistees, which accounted for about 70 percent of the entire US military personnel engaged in WWII (Selective Service System 1948; Rose 2018), missing information for the Navy, Marines, and Coast Guard. Second, the coverage of enlistment records is uneven across US states. Third, stemming from record linkage, this WWII veteran measure necessarily misses all the individuals who served in the US Army but could not be linked. Such incomplete measurement leads us to underestimate the change in probability to serve in WWII across birth cohorts. We discuss the implications for our results in detail in section 5.

3.2. College Outcomes

Our primary data source is the first release of the full count 1950 census, made available by IPUMS on January 2024 (Ruggles et al. 2021). In addition to basic demographic information such as age, birth location, and ethnicity, the 1950 census provides data on education and WWII veteran status. Since in 1950 these questions were only asked to 20 percent of the population (every fifth person on the census form), we limit our sample to these "sample-line" individuals. We obtain parental income, education, and occupation information by linking individuals in the 1950 census to their childhood records in the 1940 census, using the Multigenerational Longitudinal Panel (MLP) links (Ruggles et al. 2019).[9]

Because several of our focal birth cohorts had not completed their higher education degrees by 1950, our main outcome is an indicator for individuals who attended at least one year of

[9]The MLP linkage procedure relies on name similarity and corresponding birth location and birth year information, matching approximately 55 million individuals from the 1950 census (out of a total of 152 million) to the 1940 census.

college. To study college graduation outcomes, we use a dataset combining three 1 percent samples of the 1970 census (Ruggles et al. 2021).

4. Empirical Strategy

As discussed in section 2, WWII veterans constituted a selected subset of the US population. Therefore, a simple comparison of veterans and non-veterans would produce biased estimates of the effects of the GI Bill benefits. However, the probability of serving in WWII varied substantially across birth cohorts due to quasi-random factors related to the end of the war. Our empirical strategy leverages this variation to study the effects of the GI Bill college subsidies (and military service) with a Regression Discontinuity Design (RDD).

Figure 1 (panel (a)) shows the share of WWII veterans by birth year. Almost 80 percent of US men born between 1922 and 1926 served in WWII. Due to the reduced needs of the military as the war ended, the proportion starts decreasing for the 1927 birth cohort, dropping to 40 percent for those born in 1928 and to less than 10 percent for the 1930 cohort.

Similar to Fetter (2013), we adopt a fuzzy RDD approach, where the running variable is an individual's birth year and the cutoff is set after the last cohort of "broad" WWII conscription, between 1927 and 1928. Intuitively, we compare individuals finishing high-school around the same time and born in the vicinity of the cutoff: those born before 1928, highly likely to serve in WWII and to be eligible for GI Bill education subsidies, and those born in or after 1928, with a much lower risk to serve in WWII and be allowed to request GI Bill college aid.

More precisely, the birth year cutoff serves as an instrument for WWII veteran status in a two-stage least squares (2SLS) estimation procedure. Our first stage and reduced form regressions are specified as follows:

$$Veteran_i = \alpha_1 + \beta_1 \mathbb{1}(birth_i < c) + \gamma_1 (birth_i - c)\mathbb{1}(birth_i < c) + \delta_1 (birth_i - c)\mathbb{1}(birth_i > c) + \lambda_1 X_i + \varepsilon_{1,i} \quad (1)$$

$$y_i = \alpha_2 + \beta_2 \mathbb{1}(birth_i < c) + \gamma_2(birth_i - c)\mathbb{1}(birth_i < c) + \delta_2(birth_i - c)\mathbb{1}(birth_i > c)$$
$$+ \lambda_2 X_i + \varepsilon_{2,i} \tag{2}$$

where $Veteran_i$ is an indicator for WWII veterans and y_i is a college or inventive outcome for individual i. The indicator $\mathbb{1}(birth_i < c)$ is set to 1 if individual i was born before the cutoff c. The terms $(birth_i - c)\mathbb{1}(birth_i < c)$ and $(birth_i - c)\mathbb{1}(birth_i > c)$ denote separate linear trends for the period before and after the cutoff. X_i is a vector of controls including birth state dummies and an indicator for non-whites. The first stage estimates the discontinuity around the cutoff in WWII veteran status. The reduced form estimates the discontinuity around the cutoff for each of our inventor or college outcomes.

We obtain 2SLS estimates by estimating the following second stage specification:

$$y_i = \alpha_3 + \beta_3 \widehat{Veteran}_i + \gamma_3(birth_i - c)\mathbb{1}(birth_i < c) + \delta_3(birth_i - c)\mathbb{1}(birth_i > c)$$
$$+ \lambda_3 X_i + \varepsilon_{3,i} \tag{3}$$

where $\widehat{Veteran}_i$ is the indicator for WWII veterans instrumented with $\mathbb{1}(birth_i < c)$. Our estimate of interest is β_3, which measures the joint effect of eligibility to the GI Bill college benefits and military service on our inventor and college outcomes. More precisely, β_3 estimates the Local Average Treatment Effect (LATE) of eligibility to college aid and of military service on our outcomes of interest.

Regarding inference, we follow Kolesár and Rothe (2018) and calculate heteroskedasticity-robust standard errors. Similar to Fetter (2013), our baseline bandwidth includes the birth cohorts from 1925 to 1930. When using the 1970 census, we group individuals by year and quarter of birth. Due to the lack of quarter of birth information in the 1940 and 1950 census,

the rest of our estimates group individuals only by birth year. It should be noted that birth year information from both the 1940 and 1950 census is subject to measurement error. Rather than directly asking the respondents for their year of birth, both censuses asked their age since their last birthday. Because both censuses were administered on April 1st, anyone born after the first quarter in each year is assigned a birth year larger by one year.

Such measurement error is likely to introduce a conservative bias on our estimates. When using birth information from the 1940 and 1950 census, our cutoff is practically set between the first and second quarter of birth of 1927, rather than between 1927 and 1928, the former being a less pronounced discontinuity in WWII service. Figure 1 illustrates this graphically. Panel (b) plots the share of WWII veterans by birth year using information from the 1950 census. There's still a substantial decrease in the proportion of WWII veterans between 1927 and 1928, but the decline is less sharp compared panel (a), where birth year is measured correctly using the 1970 census.

We assess the robustness of our estimates against this measurement error in two ways. First, we re-estimate our models excluding the 1928 birth cohort, which in the 1940 and 1950 census contains individuals born both in 1927 and in 1928. Second, for the analysis of inventor outcomes, we provide estimates substituting the inventor's birth year observed in the census with the birth year recorded during our birth information search. In future updates of our paper, we plan to link to the entire 1940 and 1950 census to quarter of birth information. This will address measurement error and enable us to estimate regression models with greater granularity using quarter of birth data.

The key identification assumption of our empirical strategy is that any factor affecting college or inventive outcomes, other than higher rates of military conscription, does not systematically differ for individuals born around the cutoff. This implies that individuals' characteristics should vary smoothly around the cutoff.

Figure 2 plots several key outcomes for individuals born between 1922 and 1932. Panels

(a) and (b) show the proportion of high school graduates by birth year and by birth quarter and year, respectively. Both plots visually exclude large discontinuities around the cutoff. Nevertheless, cohorts facing high rates of WWII conscription appear to have lower rates of high school completion, likely due to the interruption of studies caused by WWII service. In panels (c) and (d) we inspect the share of individuals who completed 10th grade, too young to be eligible to serve in WWII. While there's an increasing trend starting right after 1928, we do not detect sharp discontinuities around the birth year cutoff.

Panel (e) focuses on the share of individuals who resided in a urban location in 1940, showing it does not vary discontinuously around the birth year cutoff. In panel (f), we consider the distance between an individual's residence and the closest higher education institution. Specifically, we calculate the miles separating the latitude and longitude coordinates of an individuals' residence in the 1940 census (data from the Census Place Project; Berkes et al. 2023) and the coordinates of the closest four-year university or junior college founded prior to 1940 (data from the Integrated Post Secondary Education System (IPEDS)). Again, individuals born around the cutoff appear to be highly comparable, as the average distance from higher education institutions does not vary discontinuously around the cutoff. Panel (g) shows the average value of father's annual income in 1940, which also varies smoothly around the cutoff.[10]

We also inspect the probability to link an inventor to the 1940 census by birth year. Panel (h) in Figure 2 shows similar proportions of inventors linked across cohorts, excluding any discontinuity in linkage rates for inventors born around the cutoff.

Lastly, we consider the potential influence of the Korean War's GI Bill on our estimates. Korean War veterans were eligible to a college subsidy program similar to the one for WWII veterans, though slightly less generous.[11] Panel (d) in Figure 1 shows the share of both WWII

[10]Since the 1940 census misses income information for farmers and the self-employed, we follow Collins and Wanamaker (2022) and Abramitzky et al. (2021b) and impute income to those occupation categories using information from the 1960 census.

[11]The main difference was the amount of the subsidy. Korean War veterans received a single scholarship amount to cover both tuition and living expenses, rather than a tuition allowance and a separate monthly stipend like WWII veterans.

and Korean War veterans by birth year and birth quarter. There is a substantial decrease in the likelihood of being a veteran around the birth year cutoff, driven by the end of WWII. The proportion of veterans is approximately 80 percent for the 1922-1926 cohorts, slightly decreases to 75 percent for 1927, and drops to around 63 percent for all cohorts from 1928 onwards. The Korean War GI Bill might introduce an upward bias on our estimates if its college subsidy uptake was higher for the last cohorts in our sample. We argue that this is unlikely, due to the limited size of our bandwidth, which, after the cutoff, includes only the cohorts from 1928 to 1930, close enough to face similar incentives and constraints to pursue a college degree upon returning to civilian life. Appendix Figure C1 supports this view, showing that the share of WWII and Korean War veterans who attended at least one year of college or obtained a degree is very similar among birth cohorts within our bandwidth, particularly those between 1928 and 1930. Additionally, the smooth variation of both proportion around the birth year cutoff further supports the main identification assumption of our empirical strategy.

In summary, the decrease in military service between the 1927 and 1928 birth cohorts due to the end of WWII generates quasi-random variation in eligibility to GI Bill college subsidies, which we exploit for a fuzzy RDD.

5. The Effect of the GI Bill and WWII Service on Inventor Outcomes

5.1. Effects on the Full Population and by Race

Figure 3 shows the share of inventors by birth cohort, plotting separate linear trends for cohorts before and after the birth year cutoff, both for our baseline bandwidth (solid blue lines) and for all cohorts shown in the plot (dashed gray lines). In panel (a) we focus on all inventors. We notice a decline in their proportion around the 1928 birth year cutoff. In particular, the inventor share is similar across the 1922-1927 cohorts, displaying a flat trend. It then drops to lower levels for the 1928 and 1929, further decreasing for the 1930 to 1932 birth years. In panel (b) we focus on the share of inventors who filed their first patent by 40 years old, the majority of whom

started their patenting career close to their undergraduate or graduate training. This group is likely to capture mostly "professional" inventors who started patenting at the beginning of a career as a scientist or engineer (rather than one-off independent inventors), possibly accessible only through a higher education degree. The proportion of inventors from this group increases steadily until the 1927 cohort. It then decreases and maintains a consistently lower value for the rest of birth cohorts, with a slightly decreasing trend. In panel (c) we exclude inventors patenting under their own name ("independent inventors"), tracking only inventors patenting for a company or for a government agency.[12] The plot is similar to panel (c). In panel (d) we restrict our sample to inventors patenting impactful technologies, defined as those in the top quartile of total patent citations received within their birth cohort.[13] The inventor share follows a flat trend for the 1922-1927 cohorts, dropping to consistently lower and slightly decreasing proportions for the cohorts after the birth year cutoff. Overall, we find evidence of a discontinuity around the 1928 birth year cutoff. This suggests that inventors born right before the cutoff have a higher probability to be an inventor than those born right after it.

In Table 1 we report the results of our estimates, where the dependent variable is an indicator equal to 1 for individuals who are inventors. We start by a simple comparison of WWII veterans and non-veterans (column 1), finding that the former have a substantially larger likelihood to be an inventor than those who did not serve. Column (2) reports the result of our first stage (equation (1)). The positive and statistically significant coefficient suggests that $\mathbb{1}(birth_i < c)$ is a relevant instrument for veteran status. Column (3) reports our reduced form estimate (equation (2)). The coefficient for $\mathbb{1}(birth_i < c)$ is positive and statistically significant (although only at the 10 percent level, p-value = 0.084), indicating a discontinuity in the probability to be an inventor around the birth year cutoff. In particular, our estimate implies a 0.019 percentage

[12] Out of our sample of 39,287 inventors born between 1922 and 1932 and linked to the 1940 census ("linked sample (a)"), we identify 6,654 "independent inventors", only 380 inventors patenting for an academic institution, and 2,230 inventors patenting for a government agency (based on information on federal agencies' patents from Gross and Sampat 2024).

[13] We use citations extracted from the patent's full text by Verluise et al. (2022). To address truncation for the last patent cohorts in our sample, we restrict citation counts to those recorded up to 30 years after the focal patent was filed.

points increase in the probability to become an inventor for individuals born right before the birth year cutoff, approximately 5 percent of the mean of the dependent variable.

In column (4) we report the results of our second stage (equation (3)). The estimated coefficient for $\widehat{Veteran}_i$ is positive and statistically significant (still at the 10 percent level) and implies that eligibility to GI Bill college subsidies through WWII service positively increased the probability to be an inventor. The size of the coefficient indicates a very large effect (more than three times the dependent variable's mean). This is due to mismeasurement in our veteran status variable. As discussed in section 3.1, we are able to identify only a fraction of veterans in our sample. This leads us to underestimate the size of the first stage coefficient, which in turn biases the second stage estimate (i.e., the ratio of reduced form and first stage). A back-of-the-envelope calculation where we scale the reduced form coefficient from Table 1 (column 3) by the first stage coefficient estimated in section 6.1 using the 1950 census (Table 5, panel A, column 2) returns a more credible second stage estimate, accounting for approximately 30 percent of the dependent variable's mean.[14] In light of this, we will present only estimates from our reduced form specification for the next regression exercises.

In columns (5) and (6), we re-run our reduced form regressions by race. When we focus only on white individuals, we estimate a coefficient comparable to the full sample (column 5). When we focus only on black individuals we estimate a coefficient statistically indistinguishable from zero (column 6). This is consistent with evidence showing that black individuals, particularly in the South, faced limited access to colleges, thereby limiting their ability to benefit from the GI Bill subsidies (Turner and Bound 2003).

Next, we examine how our effects vary across different inventor categories. Table 2 presents reduced form estimates where the dependent variable is restricted to a specific inventor group. In column (1) the dependent variable focuses only on inventors who filed their first patent by

[14]Specifically, the ratio of the reduced form coefficient for the probability to become an inventor (0.00019; Table 1, column 3) and the first stage coefficient estimated in the 1950 census (0.152, Table 5, panel A, column 2) is equal to 0.00125.

40 years old. We estimate a positive and statistically significant coefficient (at the 10 percent level), implying a slightly larger effect than for all inventors (approximately 7.5 percent of the dependent variable, versus 5 percent). In column (2), we exclude independent inventors, tracking only inventors patenting for companies and few others for a government agency. The estimated coefficient is positive and statistically significant (p-value = 0.037), indicating a 0.020 percentage points increase in the probability to become an inventor for cohorts just before the birth year cutoff, around 7 percent of the dependent variable's mean. In column (3), we focus only on independent inventors and we estimate a coefficient statistically indistinguishable from zero. This result suggests that GI Bill eligibility (and WWII service) increased the probability to become an inventor only for those likely to be employed in R&D laboratories. This is relevant in our context, because most of those careers typically require at least undergraduate training.

In columns (4) to (6), our dependent variable includes only inventors patenting impactful technologies. Column (4) focuses on inventors in the top quartile of total patent citations received within their birth cohort, showing a positive and statistically significant coefficient (p-value = 0.013). Column (5) tracks only inventors in the top quartile of average citations received, while column (6) tracks only those in the top quartile of the highest citations for a single patent. Both estimated coefficients are positive and statistically significant at the 1 percent level. The effect implied by these estimates ranges from 14 and 26 percent of the dependent variable's mean. These results suggest that the increase in the probability to become an inventor for cohorts eligible for GI Bill subsidies was not driven by inventors patenting low-quality technologies.

We also study the heterogeneity of our effects by inventors' technology field. Table 3 reports reduced form estimates for regressions where the dependent variable is an indicator for individuals who patented in each of six technology categories (Hall et al. 2001). We find a positive and statistically significant coefficient only for the "Electrical and electronics" category (p-value = 0.022) and for the "Computers and communications" category (significant at the 10 percent

level, p-value = 0.093). The point estimate for the first group is approximately 13 percent of the mean of the dependent variable, while that for the second group is about 5 percent. Similar to our estimates excluding independent inventors, this is an important result for our study. The technologies patented in those fields are often the result of systematic R&D by scientists and engineers in private companies' or federal laboratories, careers typically accessed only through undergraduate or graduate training. They are also fields which experienced large investments by the US government both during WWII and the postwar period, and which grew substantially in the second half of the twentieth century (Mowery and Rosenberg 1991; Mowery 2010; Gross and Sampat 2023). We further investigate inventors' sorting across technology fields in section 7.

In Appendix Figure D1 we report the coefficients for reduced form estimates where the dependent variable is constructed using the inventor-census linked sample (b) or (c). Our goal is to test the sensitivity of our estimates to a dependent variable constructed with different linkage procedures. Using the latter we find estimates very close to our baseline, with a similar coefficient size and statistical significance (p-value = 0.055). Using linked sample (b) all estimates are statistically indistinguishable from zero. We ascribe this to the underestimation of inventors in the last cohorts in our bandwidth which we discussed in section 3.1, likely biasing the estimate of the trends for cohorts above the cutoff.

We also test alternative specifications, first, by enlarging our bandwidth to 1924-1931, then, by excluding the 1928 birth cohort, and lastly by using the inventor birth year found in our birth information dataset rather than the 1940 census one. In the first case, our goal is to test the robustness of our results to a different bandwidth. In the second case, we aim to address the birth year mismeasurement we described in section 4: excluding 1928 enables us to test a sharper discontinuity in veteran status, removing a birth year which effectively includes both people born in 1927 and in 1928. In the third case, we also aim to address mismeasurement, using correct information on the inventors' birth year.

Appendix Figure D2 reports reduced form results. When the dependent variable includes all inventors in our sample, the estimates excluding the 1928 cohort and those using the inventors' birth year report slightly larger point estimates and smaller confidence intervals, relative to our baseline (p-values equal to 0.029 and equal to 0.013, respectively). The opposite is true for those based on a larger bandwidth, which produce a smaller coefficient and larger confidence intervals (p-value = 0.130). We find the same results when focusing only on white individuals, while all estimates are statistically significant when the dependent variable is restricted to inventors first patenting by 40 years old. When we use the inventors' birth year we find a positive and statistically significant coefficient also when we restrict our sample to black individuals.

5.2. Effects by Parental Background

We continue our investigation by studying how the effects on the probability of becoming an inventor vary by parental background. We consider father's education, occupation category, and income.[15] Our parental background measures are recorded when individuals in our birth cohorts bandwidth (1925-1930) were between 10 and 15 years old.

Table 4 reports reduced form estimates from our baseline specification. When we consider father's education, we find a positive and statistically significant effect only for the children of college educated individuals (panel A). In panel B we focus on five father's occupation categories: (1) farmers and farm laborers; (2) laborers, operatives, and service workers; (3) craftsmen, and clerical and sales workers; (4) managers, officials, and proprietors; and (5) professional and technical workers. We find a positive and statistically significant effect for children of "craftsmen, and clerical and sales workers" and for children of "professional and technical workers". The implied effect accounts for around 9.4 percent of the dependent variable for the first group and 20 percent for the second.

In panel C we group individuals by father's annual income quintile, while in panel D we

[15] We focus only on fathers' occupation and income outcomes due to the low female labor force participation in 1940, which was around 27.9 percent, compared to 82.5 percent for males (Blau et al. 2002; Acemoglu et al. 2004).

report estimates for five income groups: those below 1,000 dollars (approximately the median income in 1940), and four other 1,000 dollars groups. In both cases our estimates lack statistical significance to enable a conclusive interpretation. Although we find a positive coefficient for individuals in the fourth and fifth quintile, as well as for those in the 1,000-2,000 and 2,000-3,000 dollars brackets, the uncertainty around each of these estimates is too large to enable conclusive interpretations.

Figure 4 reports the reduced form coefficients from alternative specifications. Similar to our baseline estimates, when we extend our bandwidth to 1924-1931 and when we exclude the 1928 birth cohort, we find a positive and statistically significant coefficient for the children of college educated individuals, the children of "professionals and technical workers", and those of "craftsmen, and clerical and sales workers". However, when we exclude the 1928 birth cohort or when we use the inventors' birth year, we find a positive and statistically significant also for the children of individuals whose highest degree was a middle school one. Only in specifications using the inventor's birth year, we find a positive and statistically significant coefficient for estimates focusing on individuals in the fifth quintile of the father's income distribution or those within the 1,000-2,000 dollars income bracket.

A joint reading of the estimates across all our specifications in Table 4 and Figure 4 suggests that the GI Bill benefits increased access to inventor careers for individuals from the lower- to the upper-middle class. While most children from the upper middle class may have gone to college even without the GI Bill benefits, several of those born in families with lower socio-economic status may have not.

6. The Role of College Education via the GI Bill to Access Inventor Careers

Our fuzzy regressions discontinuity approach identifies the combined effect of eligibility to GI Bill college aid and military service on inventor outcomes. In this section, we investigate the individual role of college education through the GI Bill subsidies in increasing access to inven-

tor careers. First, we study the effects of the GI Bill and WWII military service on college outcomes. Second, we analyze the impact of distance from higher education institutions on inventor outcomes.

6.1. The Effect of the GI Bill and WWII Service on College Outcomes

Figure 5 (panel (a)) plots the share of individuals who attended at least one year of college, using the 1950 census.[16] The proportion increases until the 1927 cohort, dropping to slightly smaller values for cohorts between 1928 and 1930. Panel (b) depicts the share of individuals who graduated from a four-year college degree, using more granular data by birth year and quarter from the 1970 census. Overall, there's an increasing trend in the proportion of college graduates, reflecting the growing proportion of post-secondary-educated individuals among those born in the first half of the 20th century (Goldin and Katz 2010). Despite displaying a similar slope, there is a downward shift in the trend for cohorts born from 1928 onwards. Both plots show a discontinuity around the 1928 birth year cutoff.

Table 5 (panel A) reports the results of our estimates for college attendance based on the 1950 census. Our dependent variable is an indicator equal to 1 for individuals who attended at least one year of college. In column (1) we regress our dependent variable on an indicator for WWII veterans, showing that veterans are more likely to have attended college than those who did not serve in WWII. Column (2) reports the result of our first stage, showing that the birth year cutoff is a relevant instrument for WWII veteran status also in this sample. Column (3) reports our reduced form estimate. The coefficient for $\mathbb{1}(birth_i < c)$ is positive and statistically significant, indicating a discontinuity in the probability to attend college around the birth year cutoff.

Column (4) shows a positive and statistically significant estimate for our second stage, implying that individuals eligible to the GI Bill college aid through WWII service had higher rates of college attendance than those who did not have access to the program. Our estimate

[16] We show only proportions until the 1930 cohort due to unreliable college outcome information for later cohorts, too young to have started college by April 1st, 1950.

is economically significant, indicating an increase in college attendance by 6 percentage points, approximately one third of our dependent variable. In columns (5) and (6), we re-run our second stage regressions by race. When we focus only on whites, we estimate a positive statistically significant coefficient, almost identical to our estimate for the full sample. When we focus only on blacks, our coefficient is positive although imprecisely estimated (p-value = 0.161).

In panel B, we run the same estimation procedure using quarter of birth data from the 1970 census, where our dependent variable is an indicator equal to 1 for individuals who graduated from college. We find results highly comparable to those in panel A. The second stage estimate (column (4)) indicates an increase in college graduation of approximately 4 percentage points for individuals who had access to the GI Bill college aid, an effect corresponding to approximately one fourth of the dependent variable's mean.[17]

Next, we investigate the heterogeneity of the effects on college attendance by parental background, using the set of "sample-line" individuals from the 1950 linked to the 1940 census. Table 6 reports IV estimates from our baseline specification. When considering father's education (panel A), we find positive and statistically significant coefficients only for children of middle school- and high school-educated individuals. The estimate for the middle school group implies an increase in college attendance of 3.7 percentage points, around 28 percent of the dependent variable's mean, while the estimate for the high school group an increase of 14.8 percentage points, 52 percent of the dependent variable's mean.

When we focus on father's occupation (panel B), we find a positive and statistically significant estimate only for children of "craftsmen, and clerical and sales workers," indicating an increase in college attendance of 9.2 percentage points (approximately 38.6 percent of the

[17] Using a similar identification strategy, Bound and Turner (2002) find a 4 percentage point increase in college completion. Stanley (2003) employs a within-cohort empirical approach, estimating an effect ranging between 4.5 and 7.2 percentage points. Similar to our inventor analysis in section 5, we re-run the estimates in Panel A (Table 5) including the 1924 cohort to test a larger bandwidth and excluding individuals born in 1928 to address birth year measurement error (Appendix Figure D4). Overall, we find comparable results, with two differences only when we drop the 1928 birth cohort. First, the estimated coefficients for are smaller, around half of our baseline estimates. Second, we estimate a positive and statistically significant coefficient also when we restrict our sample to black individuals.

dependent variable's mean). The coefficients for the other categories are also positive, although imprecisely estimated (for children of laborers, operatives, and service workers the estimate's p-value is equal to 0.110, while for children of professional and technical workers the estimate's p-value is equal to 0.191).

In panel C we group individuals by father's annual income quintile. We find a positive and statistically significant coefficient only for the fourth and fifth quintiles. For children of individuals in the fourth quintile the estimate implies an increase in the probability to attend college of 12 percentage points (about 60 percent of the dependent variable's mean), while for those in the fifth quintile it indicates an increase by 11 percentage points (around 31 percent of the outcome's mean). It is useful to consider the specific values denoting the two quintiles. The fourth quintile includes children of individuals with an annual income between 1,125 and 1,664 dollars (or approximately 24,400 and 36,200 in 2023 US dollars), the fifth quintile includes anyone with an annual income above 1,664 dollars. In panel D we further asses the heterogeneity of effects by father's income, focusing on five income groups. Our estimates indicate an increase in the probability of attending college for those in the 1,000-2,000 and in 2,000-3,000 dollars group, similar in magnitude to the effects for the fourth and fifth quintile.

Figure 6 compares our baseline IV estimates with those obtained extending the birth year bandwidth and by excluding the 1928 birth cohort to address birth year mismeasurement. Overall, we find positive and statistically significant coefficients for the same subgroups of our baseline. In addition, we find evidence of an effect of the GI Bill and WWII service for the following categories. First, we estimate a statistically significant coefficient also for children of individuals with a college degree, implying an increase in college attendance of 5.3 percentage points. That accounts for around 10 percent of the dependent variable's mean, a smaller effect than for the children of the middle school and high school educated. Second, we find statistically significant coefficients also for the "laborers" occupation group, implying an effect slightly larger than our estimate for the "craftsmen" group. Third, we estimate a positive and statistically

significant also for individuals in the first and second quintile of the father's annual income distribution.

In general, our estimates from specifications removing the 1928 cohort to address birth year measurement error provide a similar picture for both college and inventor outcomes by parental education and occupation. For both outcomes, we find effects for the children of middle-school and college educated individuals, as well as for the children of individuals in the "craftsmen" and "professionals" occupation groups.

6.2. The Effect of Distance from Higher Education Institutions on Inventor Outcomes

We continue our investigation by studying how distance from higher education institutions influenced the likelihood to become an inventor. Following studies in the literature on the returns to education (e.g., Kane and Rouse 1993; Card 1995; Toivanen and Väänänen 2016), we use distance from a university or college as a proxy of the ease of access to higher education. Residing close to a university or a college may have facilitated the exploitation of the GI Bill benefits, for example, by reducing the need for relocation to pursue a degree or by exposing returning veterans to the counseling services set up by the Veterans Administration's across US colleges. If the GI Bill college aid played only a minor role in enabling access to inventive careers, we should expect minor differences in the probability to patent among veterans and between cohorts around the cutoff, based on their distance to higher education institutions.

As described in section 4, we calculate the distance in miles between the latitude and longitude coordinates of an individuals' residence in the 1940 census and those of the closest higher education institution. We consider only four-year universities and junior colleges founded prior to 1940 and we calculate three measures: distance from the closest of either institutions; distance from the closest four-year university; distance from the closest junior college.

In our first empirical exercise, we focus on the 1925-1930 birth cohorts and we regress an indicator equal to 1 for inventors on each distance measure. Importantly, we restrict the sample

to WWII veterans, aiming to hold constant military service factors that may have influenced the probability to enter a career conducive to patenting. All our regressions control for veterans' military grade and branch at the time of enlistment and for their ethnicity. We also add state of residence and birth year fixed effects and cluster standard errors by state of residence.

Table 7 reports the results of OLS estimates. Column 1 shows that our measure of distance from either a four-year university or a junior college is negatively associated with the probability of becoming an inventor for WWII veterans. More precisely, residing ten additional miles away from a higher education institution is associated with a decrease in the probability to become an inventor of 0.08 percentage points (or 10 percent of the mean of our dependent variable). To check whether this estimate simply captures differences between veterans from large economic centers and those from rural areas, in column 2 we add an indicator for individuals residing in a urban location. The estimated coefficient is smaller, but retains the same sign and statistical significance. In columns 3 and 4 we test the effect of distance exclusively from four-year universities and find estimates very similar to those in columns 1 and 2. When we focus only on junior colleges (columns 5 and 6), we still estimate negative and statistically significant coefficients, although their size is around half of those we obtain using the other distance measures. In columns 7 and 8 we test the effects of distance from the closest post office as a placebo. Reassuringly, both estimates are statistically indistinguishable from zero.

In our second test, we re-estimate the reduced form specification (equation (2)) adding one of our distance measures and its interaction with the indicator for individuals born prior to the cutoff ($\mathbb{1}(birth_i < c)$). Table 8 reports our results. Column 1 introduces distance from either four-year universities or junior colleges. The coefficient for $\mathbb{1}(birth_i < c)$ is positive and statistically significant, a larger and more precise estimate than its equivalent in Table 1 (column 3), still indicating the presence of a discontinuity in the likelihood to become an inventor around the cutoff. The coefficient for our distance measure is negative and statistically significant, with an economic magnitude similar to the estimates in Table 7. The coefficient for the interaction

of $\mathbb{1}(birth_i < c)$ and our distance measure is also negative and statistically significant, implying a smaller discontinuity in the likelihood to become an inventor for individuals residing farther away from a higher education institution. In other words, the higher probability of becoming an inventor for individuals born right before the cutoff decreases for those who resided farther away from a higher education institution. In column 2 we consider distance only from four-year universities and find highly comparable estimates. In column 3 we introduce distance from junior colleges, finding a negative coefficients for both distance and its interaction with $\mathbb{1}(birth_i < c)$, although the latter estimate is not statistically significant.

In summary, our evidence indicates that individuals eligible to GI Bill college subsidies who resided far from a higher education institution were less likely to access an inventor career, relative to those who lived closer to one. This result implies that college education (very likely accessed through the GI Bill) played an important role in determining whether WWII veterans pursued an inventor career.

7. WWII Military Service, the GI Bill, and the Direction of Inventive Careers

In this section, we delve deeper into the effects of the GI Bill and military service on inventors' sorting across technology fields. We adopt the same regression discontinuity empirical strategy, focusing only on inventors in our birth cohorts of interest rather than on the full US population. Below, we report reduced form estimates based on equation (2), where i indexes an inventor in our sample and $\mathbb{1}(birth_i < c)$ is an indicator equal to 1 for inventors born before the birth year cutoff.

We begin by examining inventors' likelihood to patent in each of the six NBER technology categories (Hall et al. 2001) for the full sample of inventors born between 1925 and 1930 (72,328 inventors). Panel (a) in Figure 7 shows estimates consistent with our results in section 5. Inventors born right before the birth year cutoff have a higher probability to patent in the

"Computers and communications" and the "Electrical and electronics" categories, an effect accounting for around 16 percent and 8 percent of the dependent variable, respectively. In panels (c) to (f), we test the effects of the GI Bill and military service on the probability to patent in those two technology categories by parental background, based on our inventor sample linked to the full count 1940 census ("linked sample (a)", 21,721 inventors). Our estimates indicate that the effect is driven by inventors whose father's highest level of education was either middle school or high school (panel (c)) or by inventors with fathers working in the "laborers", "craftsmen", or "managers" occupation categories (panel (d)). When considering father's income, we find positive and statistically significant effects almost exclusively for the "Computers and communications" category, either for the children of individuals with an annual income in the first, fourth, and fifth quintile (panel (e)), or between $1,000 and $2,000 (panel (f)).

A joint reading of these results and the estimates presented in section 5 suggests that the GI Bill (and military service) influenced the technology composition of inventor cohorts born right before the birth year cutoff, enlarging those cohorts with professionals working in the computing and electronics fields.

We continue our investigation by examining the probability of patenting in each of the eleven technology sub-categories making up the "Computers and communications" and "Electrical and electronics" groups. Panel (b) in Figure 7 reports our results. We find positive and statistically significant coefficients for seven sub-categories, namely: "Communications", "Computer hardware & software", "Information storage", "Electrical devices", "Power systems", "Semiconductor devices", and "Miscellaneous electronics". The largest effect is associated with the "Semiconductor devices" sub-category, indicating a 0.4 percentage point increase (33 percent of the dependent variable's mean) in the probability of patenting in those technologies for inventors born right before the birth year cutoff. Our estimate for the "Communications" sub-category implies the smallest effect, a 0.8 percentage point increase to patent in that sub-field,

approximately 13 percent of the dependent variable's mean).

These narrower technology domains coincide with many which received substantial R&D investments from the Department of Defense during WWII, becoming some of the fastest-growing industrial sectors in the postwar US economy, thanks to the initial WWII investments and to R&D contracts which university and corporate research laboratories continued to secure with the federal government after the war (Mowery and Rosenberg 1991; Mowery 2010; Gross and Sampat 2023).

This suggests three non-mutually exclusive mechanisms behind the concentration of inventors from GI Bill-eligible cohorts in those technology domains. First, WWII service may have exposed veterans to technologies used or experimented on by the military, for example, by spending part of their service close to or in US locations carrying out military R&D. Second, veterans may have pursued degrees in fields directly supported by federal R&D investments after the war, especially those continuing onto graduate coursework. Third, veterans may not have specialized narrowly in those fields during their university education, but may have been drawn into them by fast-growing industrial sectors in the postwar economy. We plan to report the results of tests for each of these mechanisms in future versions of the paper.

8. Summary of Results and Next Steps

In this paper, we studied the effects of the 1944 GI Bill on the probability to become an inventor. Using a fuzzy regression discontinuity design and novel linked inventor-census data, we found evidence of a positive effect of GI Bill eligibility on the probability to become an inventor. Our effects are driven by inventors patenting for companies or government agencies and active in the computing and electronics fields. We found positive effects also when focusing on inventors patenting high-impact technologies.

We provided evidence of the individual role of college education through the GI Bill subsidies in increasing access to inventor careers in two steps. First, we found a positive effect of GI Bill

eligibility on college attendance and on college completion. Second, we found that individuals eligible to GI Bill college subsidies who resided far from a higher education institution were less likely to access an inventor career, relative to those who lived closer.

We also explored the heterogeneity of our effects by parental background. For both inventor and college outcomes we found effects for the children of middle-school and college educated individuals, as well as for the children of parents employed as craftsmen or as clerical and sales workers and as professional or technical workers. This suggests that the GI Bill benefits increased access to inventor careers for individuals from the lower- to the upper-middle class.

Lastly, focusing only on inventors, we found that those born in GI Bill-eligible cohorts were disproportionately likely to patent in narrowly defined computing and electronics fields, such as semiconductors, computer hardware and software, and communication technologies.

We plan to improve our paper in three main ways. First, we aim to add quarter-of-birth information to all our datasets, which will enable us to address birth year mismeasurement and to produce RD estimates at a more granular level. Second, we plan to link the newly-released 1950 census to inventor records, which will allow us to directly test the role of college education in increasing access to inventor careers. Third, we plan to deepen our analysis in section 7 by exploring the mechanisms which might explain the concentration of inventors born in GI Bill-eligible cohorts in the computing and electronics technology fields.

References

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum, and S. Pérez (2021a): "Automated linking of historical data," *Journal of Economic Literature*, 59, 865–918.

Abramitzky, R., L. Boustan, E. Jácome, and S. Pérez (2021b): "Intergenerational mobility of immigrants in the United States over two centuries," *American Economic Review*, 111, 580–608.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102, 1832–56.

——— (2014): "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 122, 467–506.

Abramitzky, R., L. Greska, S. Peréz, J. Price, C. Schwarz, and F. Waldinger (2024a): "Climbing the Ivory Tower: How Socio-Economic Background Shapes Academia," *Mimeo*.

Abramitzky, R., J. Kowalski, S. Peréz, and J. Price (2024b): "Access to college education in the US over the last century," *Mimeo*.

Abramitzky, R. and V. Lavy (2014): "How responsive is investment in schooling to changes in redistributive policies and in returns?" *Econometrica*, 82, 1241–1272.

Acemoglu, D., D. H. Autor, and D. Lyle (2004): "Women, war, and wages: The effect of female labor supply on the wage structure at midcentury," *Journal of Political Economy*, 112, 497–551.

Agarwal, R., I. Ganguli, P. Gaulé, and G. Smith (2023): "Why US immigration matters for the global advancement of science," *Research Policy*, 52, 104659.

Agarwal, R. and P. Gaulé (2020): "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights*, 2, 409–24.

Aghion, P., U. Akcigit, A. Hyytinen, and O. Toivanen (2023): "2022 Klein lecture parental education and invention: the finnish enigma," *International Economic Review*, 64, 453–490.

Akcigit, U., J. Grigsby, and T. Nicholas (2017a): "Immigration and the Rise of American Ingenuity," *American Economic Review*, 107, 327–31.

——— (2017b): "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age," *NBER WP 23047, National Bureau of Economic Research.*

Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva (2022): "Taxation and Innovation in the Twentieth Century," *The Quarterly Journal of Economics*, 137, 329–385.

Althoff, L. and C. Szerman (2022): "The GI Bill and black-white wealth disparities," *Mimeo.*

Altschuler, G. and S. Blumin (2009): *The GI Bill: The new deal for veterans*, Oxford University Press.

Andrews, M. J. (2021): "Historical patent data: A practitioner's guide," *Journal of Economics & Management Strategy*, 30, 368–397.

——— (2023): "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges," *American Economic Journal: Economic Policy*, 15, 1–41.

Angrist, J. and A. B. Krueger (1994): "Why do World War II veterans earn more than nonveterans?" *Journal of Labor Economics*, 12, 74–97.

Angrist, J. D. and S. H. Chen (2011): "Schooling and the Vietnam-era GI Bill: Evidence from the draft lottery," *American Economic Journal: Applied Economics*, 3, 96–118.

Barr, A., L. Kawano, B. Sacerdote, W. Skimmyhorn, and M. Stevens (2021): "You Can't Handle The Truth: The Effects Of The Post-9/11 Gi Bill On Higher Education And Earnings," *NBER WP 29024, National Bureau of Economic Research.*

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019): "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *The Quarterly Journal of Economics*, 134, 647–713.

Bergeaud, A. and C. Verluise (2024): "A new dataset to study a century of innovation in Europe and in the US," *Research Policy*, 53, 104903.

Berkes, E., E. Karger, and P. Nencka (2023): "The census place project: A method for geolocating unstructured place names," *Explorations in Economic History*, 87, 101477.

Bernstein, S., R. Diamond, A. Jiranaphawiboon, T. J. McQuade, and B. Pousada (2022): "The Contribution of High-Skilled Immigrants to Innovation in the United States," *NBER WP 30797, National Bureau of Economic Research.*

Bianchi, N. and M. Giorcelli (2020): "Scientific Education and Innovation: from Technical Diplomas to University STEM Degrees," *Journal of the European Economic Association*, 18, 2608–2646.

Blau, F. D., M. A. Ferber, and A. E. Winkler (2002): *The economics of women, men, and work*, Upper Saddle River, N.J.: Prentice Hall.

Bloom, N., J. Van Reenen, and H. L. Williams (2019): "A toolkit of policies to promote innovation," *Journal of Economic Perspectives*, 33, 163–84.

Bound, J. and S. Turner (2002): "Going to war and going to college: Did World War II and the GI Bill increase educational attainment for returning veterans?" *Journal of Labor Economics*, 20, 784–815.

Bryan, K. A. and H. L. Williams (2021): "Innovation: market failures and public policies," in *Handbook of Industrial Organization*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, Elsevier, vol. 5, 281–388.

Card, D. (1995): "Using geographic variation in college proximity to estimate the return to schooling,," in *Aspects of labour market behviour: essays in honour of John Vanderkamp*, ed. by L. Chirstofides, E. Grant, and R. Swindinsky, University of Toronto Press, Toronto, Canada, 201–222.

Card, D. and A. B. Krueger (1992): "Does school quality matter? Returns to education and the characteristics of public schools in the United States," *Journal of political Economy*, 100, 1–40.

Chetty, R., D. J. Deming, and J. N. Friedman (2023): "Diversifying society's leaders? The causal effects of admission to highly selective private colleges," *NBER WP 31492, National Bureau of Economic Research*.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2020): "Income segregation and intergenerational mobility across colleges in the United States," *The Quarterly Journal of Economics*, 135, 1567–1633.

Choudhury, P., I. Ganguli, and P. Gaulé (2023): "Top talent, elite colleges, and migration: Evidence from the Indian Institutes of Technology," *Journal of Development Economics*, 164, 103120.

Cohen, W. M., R. R. Nelson, and J. P. Walsh (2000): "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)." *NBER WP 7552, National Bureau of Economic Research*.

Collins, W. J. and M. H. Wanamaker (2022): "African American intergenerational economic mobility since 1880," *American Economic Journal: Applied Economics*, 14, 84–117.

Dale, S. B. and A. B. Krueger (2002): "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables," *The Quarterly Journal of Economics*, 117, 1491–1527.

Deming, D. and S. Dynarski (2010): "College Aid," in *Targeting Investments in Children. Fighting Poverty When Resources Are Limited*, ed. by P. B. Levine and D. J. Zimmerman, University of Chicago Press.

Dittmar, J. E. and R. R. Meisenzahl (2020): "The Research University, Invention, and Industry: Evidence from German History," *Mimeo*.

Eden, M. (2023): "Quantifying racial discrimination in the 1944 GI bill," *Explorations in Economic History*, 90, 101542.

Fetter, D. K. (2013): "How do mortgage subsidies affect home ownership? Evidence from the mid-century GI Bills," *American Economic Journal: Economic Policy*, 5, 111–47.

Flynn, G. Q. (1993): *The Draft, 1940-1973*, University Press of Kansas.

Frederiksen, N. and W. B. Schrader (1950): "Adjustment to College. A Study of 10,000 Veteran and Nonveteran Students in Sixteen American Colleges." *Educational Testing Service, Princeton, N.J.*

Geiger, R. L. (2019): *American Higher Education Since World War II: A History*, Princeton University Press.

Goldin, C. and L. F. Katz (2010): *The Race Between Education and Technology*, Harvard University Press.

Goolsbee, A. (1998): "Does Government R&D Policy Mainly Benefit Scientists and Engineers?" *American Economic Review*, 88, 298–302.

Griliches, Z. (1977): "Estimating the returns to schooling: Some econometric problems," *Econometrica*, 1–22.

——— (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28, 1661–1707.

Gross, D. P. (2023): "The hidden costs of securing innovation: the manifold impacts of compulsory invention secrecy," *Management Science*, 69, 2318–2338.

Gross, D. P. and M. P. Roche (2023): "Coordinated RD programs and the creation of new industries," *HBS working paper 124-027*.

Gross, D. P. and B. N. Sampat (2023): "America, jump-started: World War II R&D and the takeoff of the US innovation system," *American Economic Review*, 113, 3323–3356.

——— (2024): "The Government Patent Register: A New Resource for Measuring US Government-Funded Patenting," *NBER WP 32136, National Bureau of Economic Research*.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001): "The NBER patent citation data file: Lessons, insights and methodological tools," *NBER WP 8498, National Bureau of Economic Research*.

Hoxby, C. M. (1997): "How the changing market structure of US higher education explains college tuition," *NBER WP 6323, National Bureau of Economic Research*.

Iaria, A., C. Schwarz, and F. Waldinger (2022): "Gender gaps in academia: Global evidence over the twentieth century," *Mimeo*.

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

Kaltenberg, M., A. B. Jaffe, and M. E. Lachman (2023): "Invention and the life course: Age differences in patenting," *Research Policy*, 52, 104629.

Kane, T. J. and C. E. Rouse (1993): "Labor market returns to two-and four-year college: is a credit a credit and do degrees matter?" *NBER WP 4268, National Bureau of Economic Research*.

Keller, M. and P. Keller (2001): *Making Harvard modern: The rise of America's university*, Oxford University Press.

Kennett, L. B. (1997): *GI: The American Soldier in World War II*, University of Oklahoma Press.

Kolesár, M. and C. Rothe (2018): "Inference in regression discontinuity designs with a discrete running variable," *American Economic Review*, 108, 2277–2304.

Lemieux, T. and D. Card (2001): "Education, earnings, and the "Canadian GI Bill"," *Canadian Journal of Economics/Revue canadienne d'économique*, 34, 313–344.

Lennon, C. (2023): "Women's educational attainment, marriage, and fertility: Evidence from the 1944 GI Bill," *Explorations in Economic History*, 90, 101539.

Li, G.-C., R. Lai, A. D'Amour, D. M. Doolin, Y. Sun, V. I. Torvik, Z. Y. Amy, and L. Fleming (2014): "Disambiguation and Co-Authorship Networks of the US Patent Inventor Database (1975–2010)," *Research Policy*, 43, 941–955.

Lochner, L. and E. Moretti (2004): "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports," *American economic review*, 94, 155–189.

Mansfield, E. (1986): "Patents and innovation: an empirical study," *Management science*, 32, 173–181.

Marco, A. C., M. Carley, S. Jackson, and A. Myers (2015): "The USPTO historical patent data files. Two centuries of innovation," *USPTO Economic Working Paper No. 2015-1.*

McCallum, A., K. Nigam, and L. H. Ungar (2000): "Efficient clustering of high-dimensional data sets with application to reference matching," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 169–178.

Mitchell, J. P. (1946): *Stanford University 1916-1941*, Stanford University Press.

Mokyr, J. (2005): "Long-Term Economic Growth and the History of Technology," in *Handbook of Economic Growth*, ed. by P. Aghion and S. N. Durlauf, Elsevier, vol. 1B, 1113–1180.

Monath, N., S. Madhavan, C. DiPietro, A. McCallum, and C. Jones (2021): "Disambiguating patent inventors, assignees, and their locations in patentsview," *Tech. Rep.*

Moser, P. (2016): "Patents and innovation in economic history," *Annual Review of Economics*, 8, 241–258.

Mowery, D. C. (2010): "Military R&D and innovation," in *Handbook of the Economics of Innovation*, ed. by B. H. Hall and N. Rosenberg, Elsevier, vol. 2, 1219–1256.

Mowery, D. C. and N. Rosenberg (1991): *Technology and the pursuit of economic growth*, Cambridge University Press.

Olson, K. W. (1973): "The GI Bill and higher education: Success and surprise," *American Quarterly*, 25, 596–610.

——— (1974): *The GI Bill, the veterans, and the colleges*, University Press of Kentucky.

Page, L. C. and J. Scott-Clayton (2016): "Improving college access in the United States: Barriers and policy responses," *Economics of Education Review*, 51, 4–22.

Pezzoni, M., F. Lissoni, and G. Tarasconi (2014): "How to Kill Inventors: Testing the Massacrator© Algorithm for Inventor Disambiguation," *Scientometrics*, 101, 477–504.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98, S71–S102.

——— (2001): "Should the government subsidize supply or demand in the market for scientists and engineers?" in *Innovation Policy and the Economy*, ed. by A. B. Jaffe, J. Lerner, and S. Stern, MIT Press, vol. 20, 221–252.

Rose, E. K. (2018): "The rise and fall of female labor force participation during World War II in the United States," *The Journal of Economic History*, 78, 673–711.

Ruggles, S., C. Fitch, R. Goeken, J. Hacker, J. Helgertz, E. Roberts, M. Sobek, K. Thompson, J. Warren, and J. Wellington (2019): "IPUMS Multigenerational Longitudinal Panel," .

Ruggles, S. C. A. F., R. Goeken, J. D. Hacker, M. A. Nelson, E. Roberts, M. Schouweiler, and M. Sobek (2021): "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]," *Minneapolis, MN: IPUMS*.

Sarada, S., M. J. Andrews, and N. L. Ziebarth (2019): "Changes in the demographics of American inventors, 1870–1940," *Explorations in Economic History*, 74, 101275.

Selective Service System (1948): "Quotas, Calls and Inductions," *Vol. 12 of Special Monographs, Washington DC:GPO*.

Stanley, M. (2003): "College education and the midcentury GI Bills," *The Quarterly Journal of Economics*, 118, 671–708.

Stouffer, S. A., E. A. Suchman, L. C. DeVinney, S. A. Star, and R. M. Williams Jr (1949): "The American soldier: Adjustment during army life. (Studies in social psychology in World War II), vol. 1," .

Toivanen, O. and L. Väänänen (2016): "Education and Invention," *Review of Economics and Statistics*, 98, 382–396.

Torvik, V. I., M. Weeber, D. R. Swanson, and N. R. Smalheiser (2005): "A probabilistic similarity metric for Medline records: A model for author name disambiguation," *Journal of the American Society for information science and technology*, 56, 140–158.

Trajtenberg, M., G. Shiff, and R. Melamed (2006): "The "names game": Harnessing inventors' patent data for economic research," *NBER WP 12479, National Bureau of Economic Research*.

Turner, S. and J. Bound (2003): "Closing the gap or widening the divide: The effects of the GI Bill and World War II on the educational outcomes of black Americans," *The Journal of Economic History*, 63, 145–177.

US Office of War Mobilization and Reconversion (1946): "The Veteran and Higher Education: A Report to the President," Tech. rep.

US The President's Commission on Veteran's' Pensions (1956): "Veterans' Benefits in the United States: A Report to the President." Tech. rep.

USPTO (2022): "The USPTO PatentsView database [dataset]," *https://patentsview.org/download/data-download-tables*.

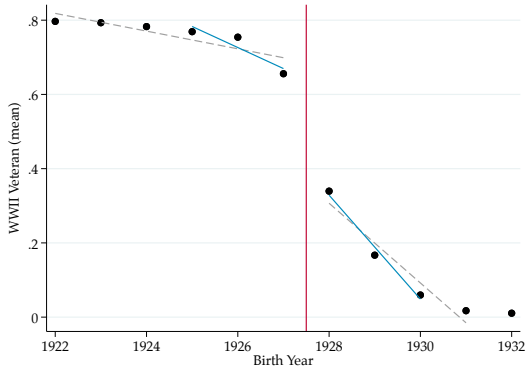
Van Reenen, J. (2022): "Innovation and Human Capital Policy," in *Innovation and Public Policy*, ed. by A. Goolsbee and B. F. Jones, University of Chicago Press, vol. 20, 61–83.

Verluise, C., G. Cristelli, K. Higham, and G. de Rassenfosse (2022): "Boosting the signal: Contextualization, extraction, and exploration of in-text patent citations," *Mimeo*.

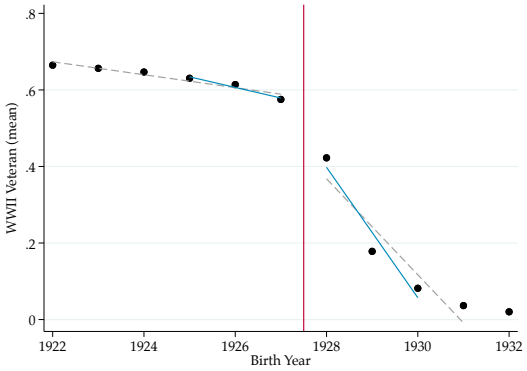
Wikle, A. and M. Osborne (2023): "CenSoc WWII Army Enlistment Dataset," *Mimeo*.

Figures

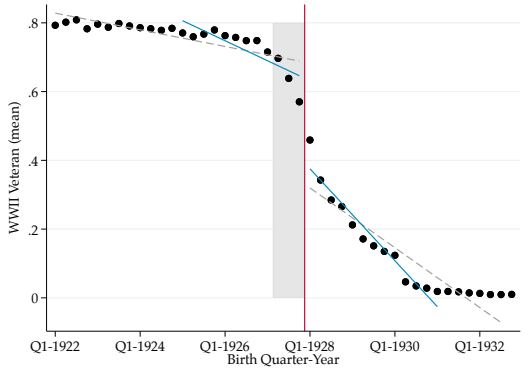
Figure 1: Veterans by birth cohort



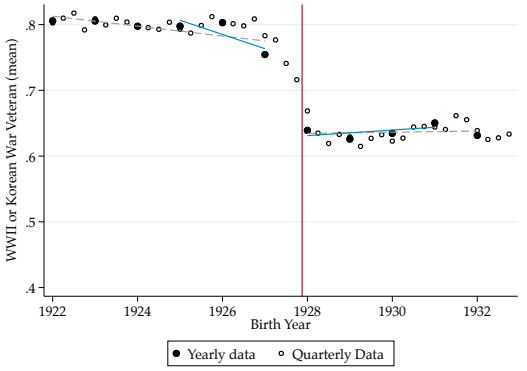
(a) WWII veterans by birth year (1970 census subsample)



(b) WWII veterans by birth year (full count 1950 census)



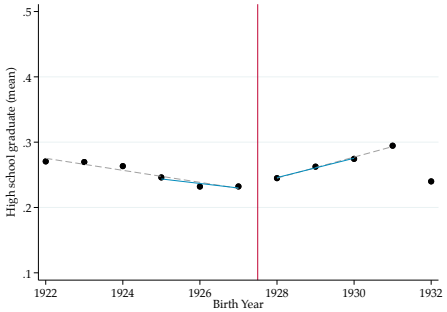
(c) WWII veterans by birth quarter and year (1970 census subsample)



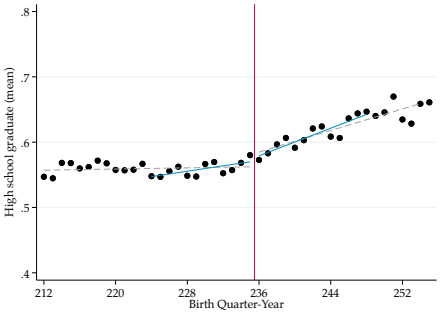
(d) WWII and Korean War veterans by birth year and by birth quarter and year (1970 census subsample)

Notes: Panel (a), (c), and (d) are based on an aggregation of three 1 percent samples from the 1970 census. Panel (a) and (c) plot the share of male US residents who served in World War II by birth year and by birth quarter and year, respectively. Panel (d) plots the share of World War II and Korean War veterans by birth year (full circles) and birth quarter and year (empty circles). Panel (b) is based on the full count 1950 census and plots the share of male US residents who served in World War II by birth year. The sample is restricted to sample line individuals (around 20 percent of the full sample), those who were asked questions about their veteran status. Dashed gray lines show linear fits for cohorts before and after the 1928 cutoff. Blue solid lines show linear fits only for the birth cohorts between 1925 and 1930. In panel (c), the shaded gray area corresponds to the second, third, and fourth birth quarters of 1927.

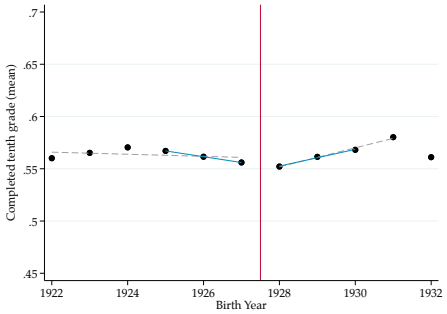
Figure 2: Individual characteristics by birth cohort



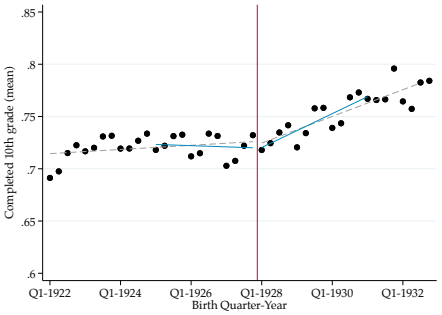
(a) High school graduates (1950 census)



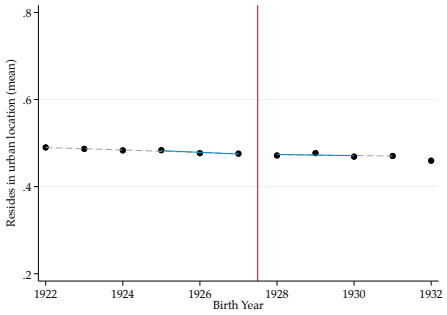
(b) High school graduates (1970 census)



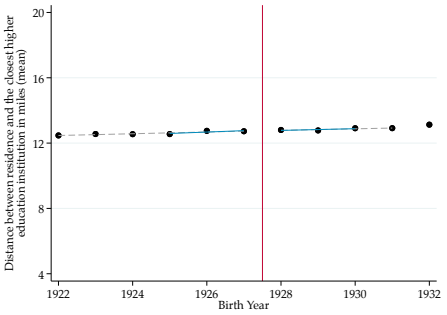
(c) Completed 10th grade (1950 census)



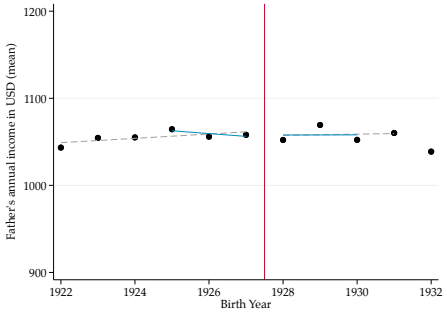
(d) Completed 10th grade (1970 census)



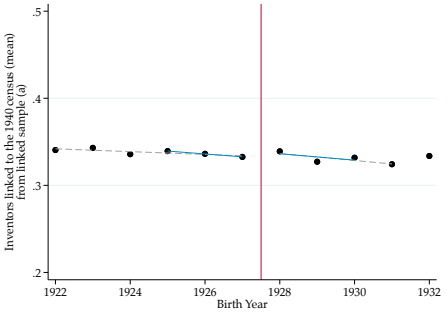
(e) Share residing in a urban location (1940 census)



(f) Distance between residence and the closest higher education institution, in miles (1940 census)



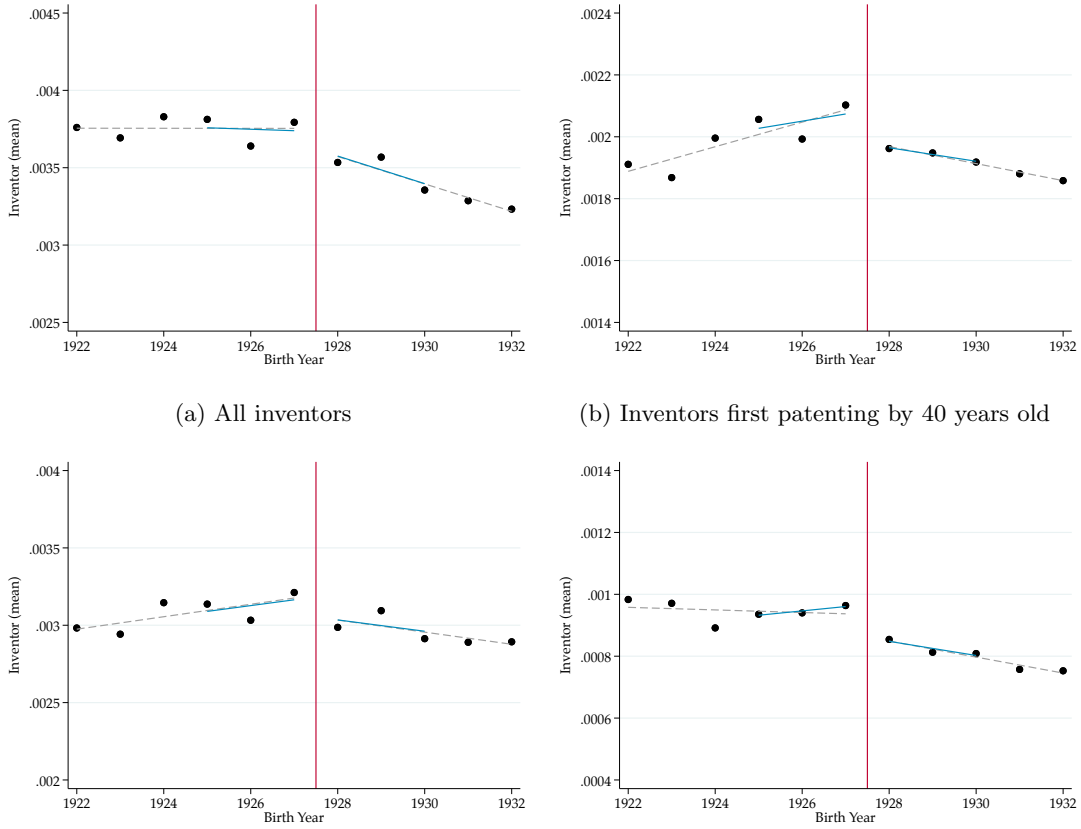
(g) Father's income (1940 census)



(h) Inventors linked to the 1940 census, linked sample (a)

Notes: Dashed gray lines show linear fits for cohorts before and after the 1928 cutoff. Blue solid lines show linear fits only for the birth cohorts between 1925 and 1930.

Figure 3: Inventors by birth cohort



(a) All inventors

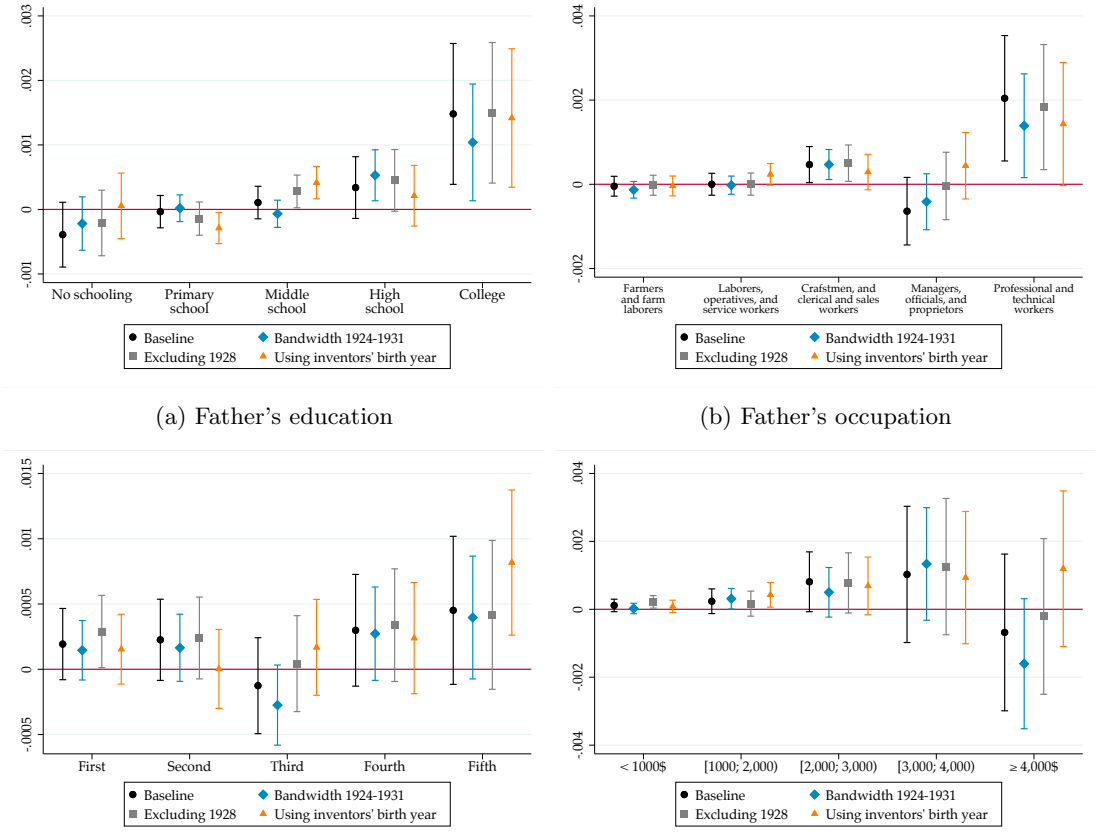
(b) Inventors first patenting by 40 years old

(c) Excluding independent inventors

(d) Top inventors by total citations

Notes: Share of inventors by birth year among US-born males from the full count 1940 census. Inventor status is based on the inventor-census linked sample (a). Panel (a) considers all inventors in the sample. Panel (b) includes only inventors who filed their first patent by 40 years old. Panel (c) excludes inventors filing patents under their own name ("independent inventors"). Panel (d) considers only inventors in the top quartile of citations received for their birth cohort. Dashed gray lines show linear fits for cohorts before and after the 1928 cutoff. Blue solid lines show linear fits only for the birth cohorts between 1925 and 1930.

Figure 4: Reduced form estimates: probability to become an inventor by parental background (alternative estimates)



(a) Father's education

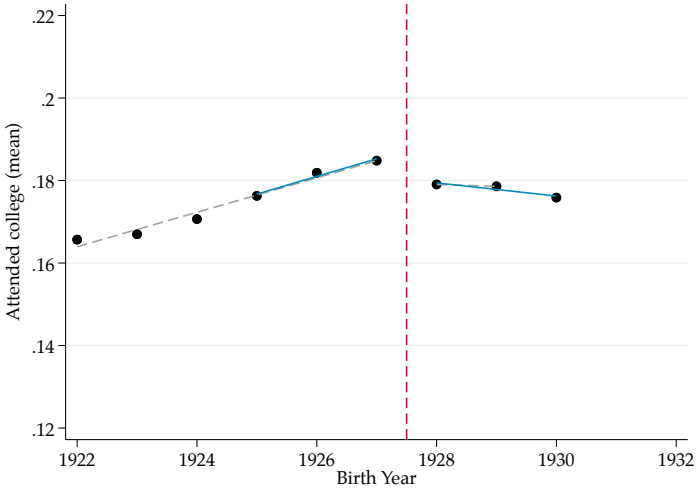
(b) Father's occupation

(c) Father's income quintile

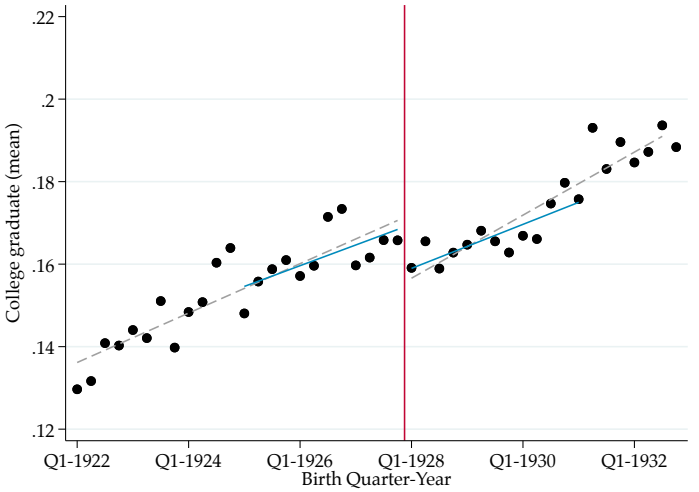
(d) Father's income category

Notes: Each coefficient plots the reduced form estimate and 90 percent confidence interval from equation (2). The dependent variable is an indicator equal to 1 for inventors (based on inventor-census linked sample (a)). Black markers report estimates from baseline regressions with the 1925-1930 birth year bandwidth. Blue markers are based on regressions where the birth year bandwidth is extended to 1924-1931. Gray markers report estimates from regressions excluding the 1928 birth year, considering the birth years 1925, 1926, and 1927 (before the cutoff) and 1929, 1930, and 1931 (after the cutoff). Orange markers report estimates where for inventors we use the birth year identified with our birth information search, instead of the birth year found in the 1940 census. All regressions include birth state dummies, an indicator for non-whites, and linear trends before and after the birth year cutoff. Estimates by OLS.

Figure 5: Individuals who attended or graduated from college by birth cohort



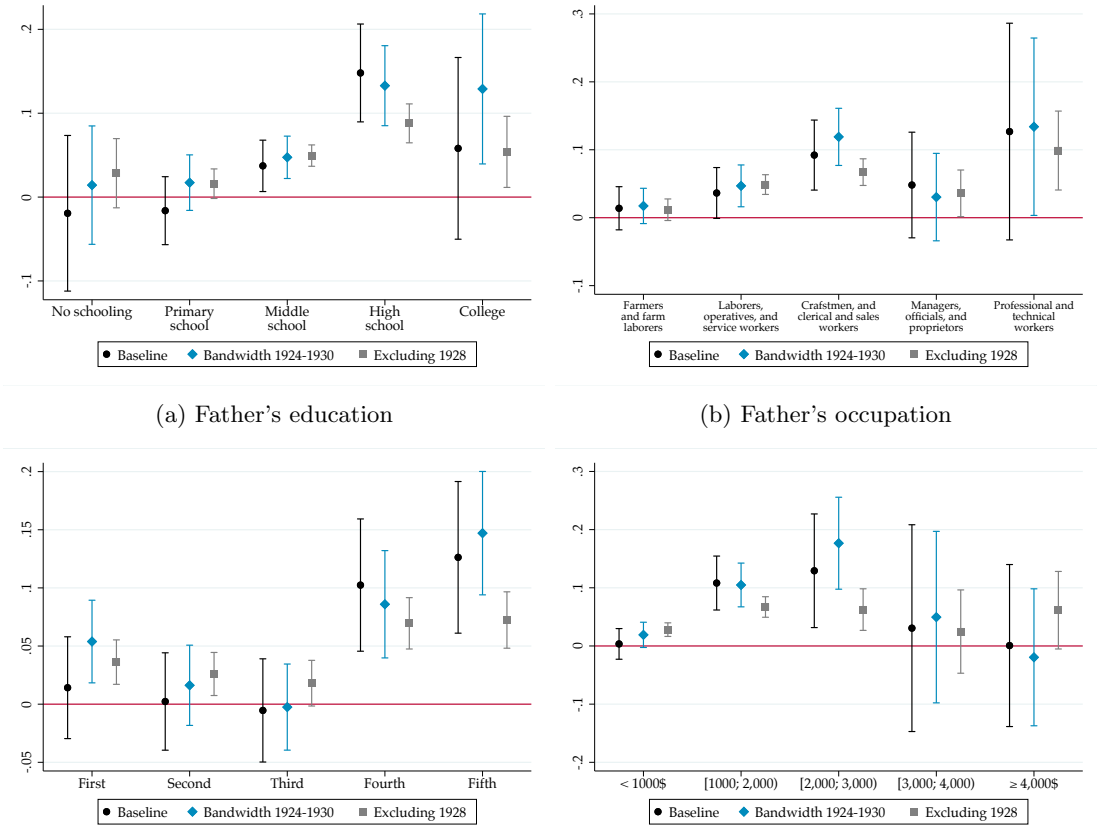
(a) Individuals who attended college by birth year
(full count 1950 Census)



(b) College graduates by birth quarter and year
(subsample of 1970 Census)

Notes: Panel (a) is based on the full count 1950 census and plots the share of male US residents who attended at least one year of college by birth year. The sample is restricted to sample lines individuals (around 20 percent of the full sample), those who were asked questions about their education level. The last birth year plotted is 1930 because that is the last cohort entering college by April 1950. Panel (b) is based on an aggregation of three 1 percent samples from the 1970 census, plotting the share of male US residents who graduated from college by birth quarter. Dashed gray lines show linear fits for cohorts before and after the 1928 cutoff. Blue solid lines show linear fits only for the birth cohorts between 1925 and 1930.

Figure 6: IV estimates for college attendance by parental background (alternative empirical specifications)



(a) Father's education

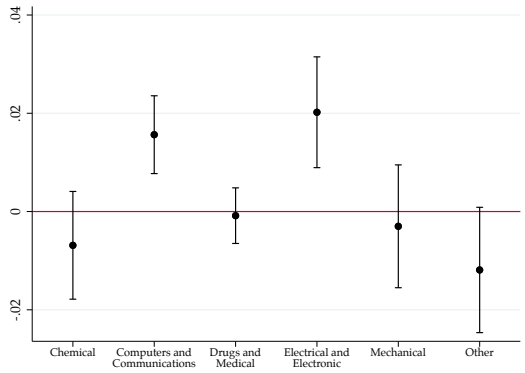
(b) Father's occupation

(c) Father's income quintile

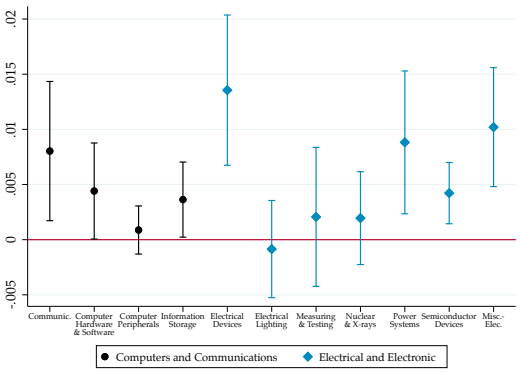
(d) Father's income category

Notes: Each coefficient plots the IV estimate and 90 percent confidence interval for the effect of veteran status on college attendance. The dependent variable is an indicator equal to 1 for individuals who attended at least one year of college. Black markers report estimates from baseline regressions with the 1925-1930 birth year bandwidth. Blue markers are based on regressions where the birth year bandwidth is extended to 1924-1931. Gray markers report estimates from regressions excluding the 1928 birth year, and considering the birth years 1925, 1926, and 1927 (before the cutoff) and 1929, 1930, and 1931 (after the cutoff). All regressions include birth state dummies, an indicator for non-whites, and linear trends before and after the birth year cutoff. Estimates by 2SLS.

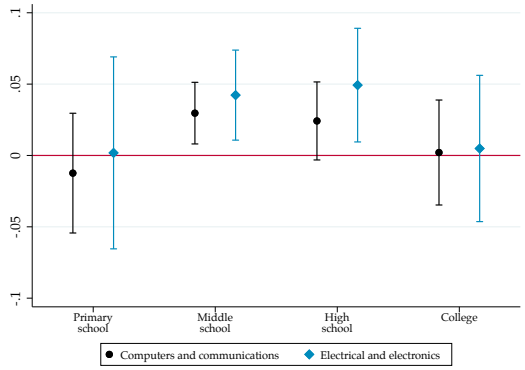
Figure 7: Reduced form (inventors only): probability to patent in a given technology category



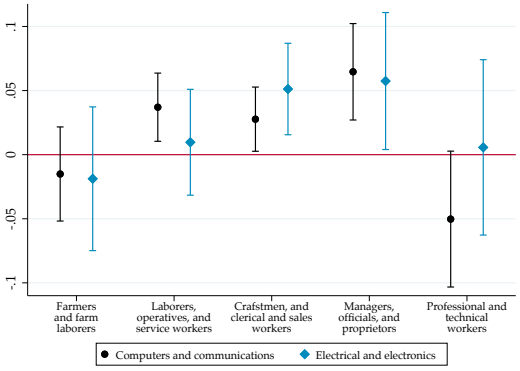
(a) By technology category



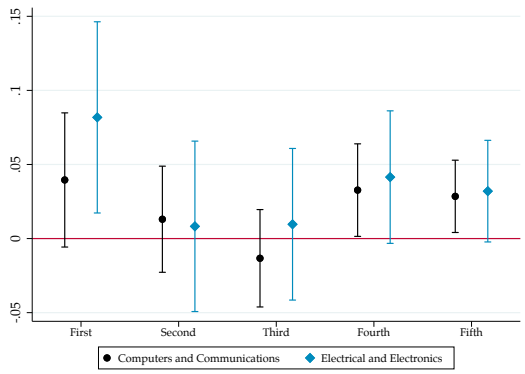
(b) By technology sub-category



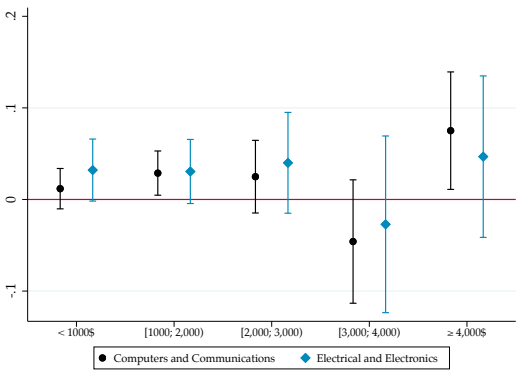
(c) By father's education and technology category



(d) By father's occupation and technology category



(e) By father's income quintile and technology category



(f) By father's income category and technology category

Notes: Estimates based on our sample of inventors born between 1925 and 1930. Each coefficient plots the reduced form estimate and 90 percent confidence interval from equation (2), where i indexes an inventor. The dependent variable is an indicator equal to 1 for inventors who patented in a specific technology category at least once during their career (based on the NBER technology categorization for patents, Hall et al. (2001)). Panels (c) to (f) focus only on the "Computers and communications" and "Electrical and electronic" categories.

Tables

Table 1: The effects of the GI Bill and WWII service on the probability to become an inventor

	OLS	First stage	Reduced form	IV	Reduced form (only whites)	Reduced form (only blacks)
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(birth_i < c)$		0.01249***	0.00019*		0.00021*	-0.000005
		(0.00050)	(0.00011)		(0.00012)	(0.00013)
$Veteran_i$	0.00463***			0.01488*		
	(0.00014)			(0.00861)		
Mean dep. var.	0.004	0.071	0.004	0.004	0.004	0.0005
Observations	6,063,629	6,063,629	6,063,629	6,063,629	5,499,036	535,432

Notes: The sample includes all men born in the US between 1925 and 1930 from the full count 1940 census. Except column (2), the dependent variable is an indicator equal to 1 for inventors (based on inventor-census linked sample (a)). In column (2), the dependent variable is an indicator equal to 1 for World War II Army and Air Force veterans (based on links to the World War II Army Enlistment Records). The birth year cutoff is 1928. All regressions include birth state dummies. Columns (2) to (6) include also linear trends before and after the birth year cutoff. Columns (2) to (4) include also an indicator for nonwhites. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS and 2SLS. * p<0.1, ** p<0.05, *** p<0.01.

Table 2: Reduced form estimates: probability to become an inventor by patenting group

	Inventor patenting by 40 years old	Excluding independent inventors	Only independent inventors	Top inventors: total citations	Top inventors: average citations	Top inventors: highest citations
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(birth_i < c)$	0.00015*	0.00020**	0.00001	0.00013**	0.00025***	0.00020***
	(0.00008)	(0.00010)	(0.00004)	(0.00005)	(0.00006)	(0.00006)
Mean dep. var.	0.002	0.0028	0.0006	0.0009	0.0010	0.010
Observations	6,063,629	6,063,629	6,063,629	6,063,629	6,063,629	6,063,629

Notes: The sample includes all men born in the US between 1925 and 1930 from the full count 1940 census. The dependent variable is an indicator equal to 1 for inventors (based on inventor-census linked sample (a)). In column (1), the dependent variable is restricted to inventors who filed their first patent by the age of 40 years old. In column (2), the dependent variable excludes inventors filing patents under their own name ("independent inventors"), thus, including only inventors patenting mostly for companies and government agencies. In column (3), the dependent variable includes only independent inventors. In column (4), the dependent variable includes only inventors in the top quartile of citations received for their birth cohort. In column (5), the dependent variable includes only inventors in the top quartile of average citations received by their patent, in the inventors' birth cohort. In column (6), the dependent variable includes only inventors in the top quartile of the highest number of citations received by a single patent, for their birth cohort. The birth year cutoff is 1928. All regressions include linear trends before and after the birth year cutoff, birth state dummies, and an indicator for nonwhites. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.

Table 3: Reduced form estimates: probability to become an inventor by technology category

	Chemical	Computers and communications	Drugs and medical	Electrical and electronic	Mechanical	Other
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(birth_i < c)$	0.00003	0.00006*	0.00001	0.00012**	0.00007	0.00008
	(0.00005)	(0.00003)	(0.00002)	(0.00005)	(0.00006)	(0.00007)
Mean dep. var.	0.0008	0.0003	0.0002	0.0009	0.0013	0.0014
Observations	6,063,629	6,063,629	6,063,629	6,063,629	6,063,629	6,063,629

Notes: The sample includes all men born in the US between 1925 and 1930 from the full count 1940 census. The dependent variable is an indicator equal to 1 for inventors who patented in a specific technology category at least once (based on inventor-census linked sample (a) and the NBER technology categorization for patents, Hall et al. (2001)). The birth year cutoff is 1928. All regressions include linear trends before and after the birth year cutoff, birth state dummies, and an indicator for nonwhites. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.

Table 4: Reduced form estimates: probability to become an inventor by parental background

	A. Father's education				
	No schooling	Primary school	Middle school	High school	College
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(birth_i < c)$	-0.00039	-0.00003	0.00011	0.00034	0.00148**
	(0.00031)	(0.00015)	(0.00015)	(0.00029)	(0.00066)
Mean dep. var.	0.001	0.002	0.003	0.005	0.009
Observations	289,262	1,280,353	2,699,756	1,233,390	417,995
	A. Father's occupation				
	Farmers and farm laborers	Laborers, operatives, and service workers	Craftsmen, and clerical and sales workers	Managers, officials, and proprietors	Professional and technical workers
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(birth_i < c)$	-0.00005	0.00000	0.00047*	-0.00064	0.00204**
	(0.00014)	(0.00016)	(0.00026)	(0.00049)	(0.00091)
Mean dep. var.	0.002	0.003	0.005	0.006	0.010
Observations	1,557,191	2,034,096	1,480,070	521,852	233,046
	A. Father's income quintile				
	First	Second	Third	Fourth	Fifth
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(birth_i < c)$	0.00019	0.00023	-0.00013	0.00030	0.00045
	(0.00017)	(0.00019)	(0.00022)	(0.00026)	(0.00034)
Mean dep. var.	0.002	0.002	0.003	0.004	0.007
Observations	1,238,655	1,201,964	1,132,581	1,184,497	1,171,372
	A. Father's income group				
	< 1000$	[1000; 2,000)	[2,000; 3,000)	[3,000; 4,000)	≥ 4,000$
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(birth_i < c)$	0.00011	0.00024	0.00081	0.00103	-0.00068
	(0.00011)	(0.00022)	(0.00054)	(0.00122)	(0.00140)
Mean dep. var.	0.002	0.004	0.007	0.010	0.012
Observations	3,458,141	1,748,164	482,672	127,054	113,038

Notes: The sample includes all men born in the US between 1925 and 1930 from the full count 1940 census. The dependent variable is an indicator equal to 1 for inventors (based on inventor-census linked sample (a)). The birth year cutoff is 1928. All regressions include birth state dummies, an indicator for non-whites, and linear trends before and after the birth year cutoff. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.

Table 5: The effects of the GI Bill and WWII service on college attendance and graduation

	A. College attendance (1950 census)					
	OLS	First stage	Reduced form	IV	IV (Whites)	IV (Blacks)
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(birth_i < c)$		0.152***	0.009***			
		(0.002)	(0.001)			
$Veteran_i$	0.040***			0.060***	0.062***	0.037
	(0.001)			(0.010)	(0.010)	(0.027)
Mean dep. var.	0.179	0.426	0.179	0.179	0.194	0.055
Observations	1,269,905	1,269,905	1,269,905	1,269,905	1,128,003	130,387
QOB specification	No	No	No	No	No	No
	B. College graduation (1970 census)					
	OLS	First stage	Reduced form	IV	IV (Whites)	IV (Blacks)
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(birth_i < c)$		0.244***	0.009***			
		(0.004)	(0.003)			
$Veteran_i$	0.008***			0.039***	0.038***	0.070
	(0.002)			(0.014)	(0.014)	(0.118)
Mean dep. var.	0.164	0.460	0.164	0.164	0.178	0.048
Observations	194,382	194,382	194,382	194,382	173,862	18,814
QOB specification	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The estimates in panel (A) are based on sample-line individuals from the 1950 census (∼20 percent of the total). Except for column (2), the dependent variable is an indicator for individuals who attended at least one year of college. In column (2), the dependent variable is an indicator for individuals who served in World War II. The estimates in panel (B) are based on three 1 percent samples from the 1970 census. Except for column (2), the dependent variable is an indicator for individuals who graduated from college. In column (2), the dependent variable is an indicator for individuals who served in World War II. The birth year cutoff is 1928. The birth quarter cutoff is the first quarter of 1928. All regressions include birth state dummies and an indicator for non-whites. Except for column (1) in both panels, all regressions include linear trends before and after the birth year or birth quarter cutoff. In panel (B), regressions also include birth quarter dummies. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS and 2SLS. * p<0.1, ** p<0.05, *** p<0.01.

Table 6: IV estimates: college attendance by parental background

	A. Father's education				
	No schooling	Primary school	Middle school	High school	College
	(1)	(2)	(3)	(4)	(5)
$Veteran_i$	-0.019	-0.016	0.037**	0.148***	0.058
	(0.056)	(0.025)	(0.019)	(0.035)	(0.066)
Mean dep. var.	0.073	0.065	0.132	0.284	0.528
Observations	21,863	104,231	254,796	115,405	37,792
	B. Father's occupation				
	Farmers and farm laborers	Laborers, operatives, and service workers	Craftsmen, and clerical and sales workers	Managers, officials, and proprietors	Professional and technical workers
	(1)	(2)	(3)	(4)	(5)
$Veteran_i$	0.014	0.036	0.092***	0.048	0.127
	(0.019)	(0.023)	(0.031)	(0.047)	(0.097)
Mean dep. var.	0.072	0.113	0.238	0.391	0.494
Observations	140,033	185,966	146,167	51,080	21,560
	C. Father's income quintile				
	First	Second	Third	Fourth	Fifth
	(1)	(2)	(3)	(4)	(5)
$Veteran_i$	0.011	0.009	-0.003	0.120***	0.110***
	(0.026)	(0.025)	(0.028)	(0.035)	(0.040)
Mean dep. var.	0.081	0.095	0.155	0.199	0.357
Observations	112,164	109,620	105,496	109,292	108,885
	D. Father's income group				
	$<$ 1000$	[1000; 2,000)	[2,000; 3,000)	[3,000; 4,000)	$\geq$ 4,000$
	(1)	(2)	(3)	(4)	(5)
$Veteran_i$	0.004	0.108***	0.129**	0.031	0.001
	(0.016)	(0.028)	(0.059)	(0.108)	(0.085)
Mean dep. var.	0.106	0.201	0.342	0.479	0.595
Observations	295,488	175,872	50,238	13,030	10,829

Notes: The sample includes all men born in the US between 1925 and 1930 from the subset of 1950 census sample-line individuals linked to the 1940 census. The dependent variable is an indicator equal to 1 for individuals who attended at least one year of college. The birth year cutoff is 1928. All regressions include birth state dummies, an indicator for non-whites, and linear trends before and after the birth year cutoff. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by 2SLS. * p<0.1, ** p<0.05, *** p<0.01.

Table 7: The effect of distance from higher education institutions on the probability to become an inventor for World War II veterans

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance-Higher-Educ-Inst	-0.00008***	-0.00005***						
	(0.00001)	(0.00001)						
Distance-Four-Year-Univ			-0.00007***	-0.00004***				
			(0.00001)	(0.00001)				
Distance-Junior-College					-0.00004***	-0.00002***		
					(0.00001)	(0.00001)		
Distance-Post-Office							0.00000	-0.00002
							(0.00005)	(0.00005)
$\mathbb{1}$(Non-white)	-0.00425***	-0.00447***	-0.00425***	-0.00448***	-0.00424***	-0.00449***	-0.00413***	-0.00445***
	(0.00068)	(0.00072)	(0.00067)	(0.00072)	(0.00066)	(0.00071)	(0.00065)	(0.00071)
$\mathbb{1}$(Urban)		0.00294***		0.00304***		0.00319***		0.00350***
		(0.00041)		(0.00040)		(0.00039)		(0.00040)
Military grade and branch dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State of residence FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Observations	398,730	398,730	398,730	398,730	398,730	398,730	398,730	398,730

Notes: The sample includes only World War II veterans born in the US between 1925 and 1930. That is, all men born in the US between 1925 and 1930 from the full count 1940 census and linked to the World War II Army Enlistment records. The dependent variable is an indicator equal to 1 for inventors (based on inventor-census linked sample (a)). "Distance-Higher-Educ-Inst" measures the distance in miles from an individual's residence and the closest higher education institution. "Distance-Four-Year-Univ" and "Distance-Junior-College" measure the distance from the closest four-year university and the closest junior college, respectively. "Distance-Post-Office" measures the distance from the closest post office. Standard errors clustered at the state of residence level are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.

Table 8: Reduced form estimates: probability to become an inventor and distance from higher education institutions

	(1)	(2)	(3)
$\mathbb{1}(birth_i < c)$	0.000312**	0.000303**	0.000269**
	(0.000121)	(0.000122)	(0.000126)
Distance-Higher-Educ-Inst	-0.000036***		-0.000022***
	(0.000002)		(0.000001)
$\mathbb{1}(birth_i < c)\times$ Distance-Higher-Educ-Inst	-0.000009***		-0.000006***
	(0.000003)		(0.000002)
Distance-Four-Year-Univ		-0.000031***	-0.000019***
		(0.000002)	(0.000001)
$\mathbb{1}(birth_i < c)\times$ Distance-Four-Year-Univ		-0.000007***	-0.000005***
		(0.000002)	(0.000002)
Distance-Junior-College			-0.000016***
			(0.000001)
$\mathbb{1}(birth_i < c)\times$ Distance-Junior-College			-0.000002
			(0.000001)
Mean dep. var.	0.004	0.004	0.004
Observations	5,755,209	5,755,209	5,755,209

Notes: The sample includes all men born in the US between 1925 and 1930 from the full count 1940 census. The dependent variable is an indicator equal to 1 for inventors (based on inventor-census linked sample (a)). "Distance-Higher-Educ-Inst" measures the distance in miles from an individual's residence and the closest higher education institution. "Distance-Four-Year-Univ" and "Distance-Junior-College" measure the distance from the closest four-year university and the closest junior college, respectively. The birth year cutoff is 1928. All regressions include linear trends before and after the birth year cutoff, birth state dummies, and an indicator for nonwhites. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.

Appendix

"The Effects of Higher Education Subsidies on Innovation: Evidence from the GI Bill"

by Ran Abramitzky and Gabriele Cristelli

September 2024

Contents

A The 1944 US Servicemen's Readjustment Act **2**

B Database Construction **3**

B.1 Historical Patent Data 3

B.2 Inventor Disambiguation 4

B.3 Inventor Birth Information Search 11

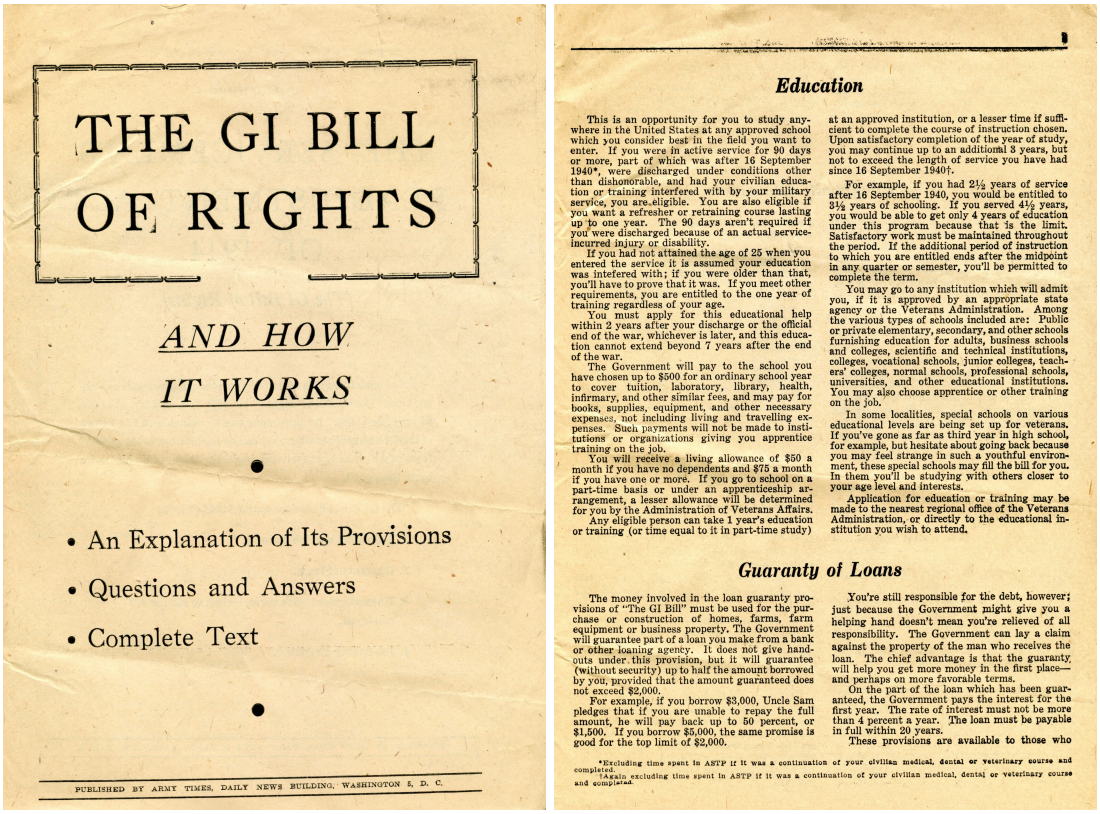
B.4 Inventor-Census Linkage 15

C Additional Descriptive Evidence **19**

D Additional Estimates and Robustness Checks **20**

A. The 1944 US Servicemen's Readjustment Act

Figure A1: GI Bill Infoset (1944)



THE GI BILL OF RIGHTS

AND HOW IT WORKS

- An Explanation of Its Provisions
- Questions and Answers
- Complete Text

PUBLISHED BY ARMY TIMES, DAILY NEWS BUILDING, WASHINGTON 5, D. C.

Education

This is an opportunity for you to study anywhere in the United States at any approved school which you consider best in the field you want to enter. If you were in active service for 90 days or more, part of which was after 16 September 1940*, were discharged under conditions other than dishonorable, and had your civilian education or training interfered with by your military service, you are eligible. You are also eligible if you want a refresher or retraining course lasting up to one year. The 90 days aren't required if you were discharged because of an actual service-incurred injury or disability.

If you had not attained the age of 25 when you entered the service it is assumed your education was intefered with; if you were older than that, you'll have to prove that it was. If you meet other requirements, you are entitled to the one year of training regardless of your age.

You must apply for this educational help within 2 years after your discharge or the official end of the war, whichever is later, and this education cannot extend beyond 7 years after the end of the war.

The Government will pay to the school you have chosen up to $500 for an ordinary school year to cover tuition, laboratory, library, health, infirmary, and other similar fees, and may pay for books, supplies, equipment, and other necessary expenses, not including living and travelling expenses. Such payments will not be made to institutions or organizations giving you apprentice training on the job.

You will receive a living allowance of $50 a month if you have no dependents and $75 a month if you have one or more. If you go to school on a part-time basis or under an apprenticeship arrangement, a lesser allowance will be determined for you by the Administration of Veterans Affairs.

Any eligible person can take 1 year's education or training (or time equal to it in part-time study) at an approved institution, or a lesser time if sufficient to complete the course of instruction chosen. Upon satisfactory completion of the year of study, you may continue up to an additional 3 years, but not to exceed the length of service you have had since 16 September 1940†.

For example, if you had 2½ years of service after 16 September 1940, you would be entitled to 3½ years of schooling. If you served 4½ years, you would be able to get only 4 years of education under this program because that is the limit. Satisfactory work must be maintained throughout the period. If the additional period of instruction to which you are entitled ends after the midpoint in any quarter or semester, you'll be permitted to complete the term.

You may go to any institution which will admit you, if it is approved by an appropriate state agency or the Veterans Administration. Among the various types of schools included are: Public or private elementary, secondary, and other schools furnishing education for adults, business schools and colleges, scientific and technical institutions, colleges, vocational schools, junior colleges, teachers' colleges, normal schools, professional schools, universities, and other educational institutions. You may also choose apprentice or other training on the job.

In some localities, special schools on various educational levels are being set up for veterans. If you've gone as far as third year in high school, for example, but hesitate about going back because you may feel strange in such a youthful environment, these special schools may fill the bill for you. In them you'll be studying with others closer to your age level and interests.

Application for education or training may be made to the nearest regional office of the Veterans Administration, or directly to the educational institution you wish to attend.

Guaranty of Loans

The money involved in the loan guaranty provisions of "The GI Bill" must be used for the purchase or construction of homes, farms, farm equipment or business property. The Government will guarantee part of a loan you make from a bank or other loaning agency. It does not give handouts under this provision, but it will guarantee (without security) up to half the amount borrowed by you, provided that the amount guaranteed does not exceed $2,000.

For example, if you borrow $3,000, Uncle Sam pledges that if you are unable to repay the full amount, he will pay back up to 50 percent, or $1,500. If you borrow $5,000, the same promise is good for the top limit of $2,000.

You're still responsible for the debt, however; just because the Government might give you a helping hand doesn't mean you're relieved of all responsibility. The Government can lay a claim against the property of the man who receives the loan. The chief advantage is that the guaranty will help you get more money in the first place—and perhaps on more favorable terms.

On the part of the loan which has been guaranteed, the Government pays the interest for the first year. The rate of interest must not be more than 4 percent a year. The loan must be payable in full within 20 years.

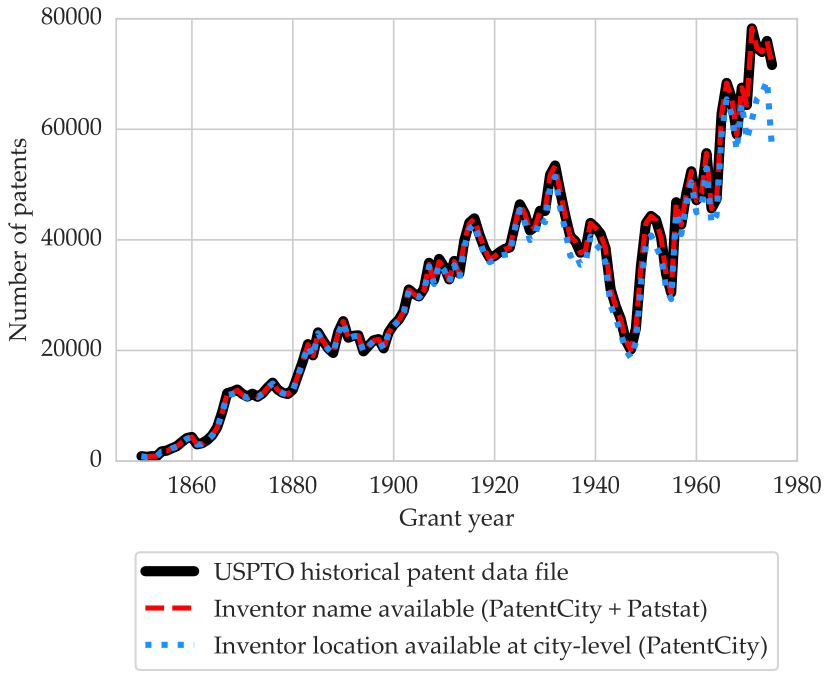
These provisions are available to those who

*Excluding time spent in ASTP if it was a continuation of your civilian medical, dental or veterinary course and completed.
†Again excluding time spent in ASTP if it was a continuation of your civilian medical, dental or veterinary course and completed.

B. Database Construction

B.1. Historical Patent Data

Figure B1: Historical USPTO patents (1850-1975)



Notes: Comparison between the total number of patents granted by the USPTO in each year between 1850-1975 (USPTO historical patent data file, Marco et al. 2015) with the number of patents with available inventors' name (PatentCity, Bergeaud and Verluise 2024) and the Patstat database, and with the number of patents with available inventors' location at the city-level (PatentCity, Bergeaud and Verluise 2024).

B.2. Inventor Disambiguation

Patent data do not natively provide unique identifiers for the same inventor across different patents. That is because intellectual property offices such as the USPTO do not require inventors to submit their information using standardized identifiers during the patent application process. Without them, researchers interested in studying individual inventors' careers face name ambiguity issues, unable to determine whether inventors sharing the same name correspond to the same person or to homonyms. Over the past two decades, scholars reliant on patent data developed disambiguation algorithms to obtain consistent inventor identifiers (e.g., Trajtenberg et al. 2006; Li et al. 2014; Pezzoni et al. 2014). These algorithms group inventors with the same name and leverage auxiliary information from patent documents, such as inventors' residential locations, co-inventors, patent assignees, and technology classes to identify inventor records referring to the same individual. In 2015, the USPTO adopted an inventor disambiguation algorithm developed by Monath et al. (2021), distributing the unique inventor identifiers generated by the algorithm in USPTO's database "PatentsView".[18]

Despite being the most comprehensive source of USTO patent data, the PatentsView database includes only "modern" patents, those granted since 1975. That does not suit the purposes of our project, which relies also on "historical" patents, those granted before 1975 (Andrews 2021). To address this challenge, we disambiguate all inventors filing a patent at the USPTO and residing in the US between 1920 and 2015. We adopt Monath et al.'s 2021 algorithm, a machine learning procedure combining supervised classification and hierarchical agglomerative clustering (HAC), and extend it to disambiguate inventors listed on both modern and historical patents.

We start by creating a training set of inventors from modern patents disambiguated by Monath et al. (2021). Similar to Akcigit et al. (2022)—who also generate unique identifiers for historical inventors using a training set of disambiguated modern inventors—we rely on two assumptions.[19] First, we assume that Monath et al.'s 2021 disambiguation is generally correct. We find this assumption reasonable: evaluations of Monath et al.'s 2021 algorithm against manually labeled inventor data reveal minimal errors, outperforming previous disambiguation algorithms, making it the best inventor disambiguation currently available to patent data researchers. Second, we assume that the features enabling the disambiguation of modern inventors have the same predictive power for the disambiguation of inventors from historical inventors.

Our training set is based on a random extraction of 2 million disambiguated inventor men-

[18]The source code for PatentsView disambiguation process based on Monath et al.'s 2021 algorithm can be found at `https://github.com/PatentsView/PatentsView-Disambiguation`.

[19]Akcigit et al. (2022) use inventor records disambiguated by Li et al. (2014).

tions (i.e., unique inventor-patent instances). We group inventor mentions into "canopies" (McCallum et al. 2000), that is, clusters of inventor mentions sharing the same last name and the first two initials of their first name. Each canopy contains a set of inventor pairs which may refer to the same individual. We obtain a training set of 25,644,056 inventor pairs. We also create a development set of 64,725 inventor pairs, based on a subsequent random extraction of 100,000 disambiguated inventor mentions not found in the training set.

Next, we train a logistic classifier to predict whether each inventor pair refers to the same individual. Following Monath et al. (2021), our features include:

- Same first name (binary)
- Jaro-Winkler similarity score for the first name
- Same middle name or no middle name for both records (binary)
- Same city of residence (binary)
- Same county of residence (binary)
- Same state of residence (binary
- Same assignee string (binary)
- Assignee string Jaro-Winkler similarity > 0.9 (binary)
- At least one common co-inventor (binary)
- Same NBER technology category, USPC technology class, and USPC technology sub-class (all binary)[20]

In order to fine-tune our algorithm, we use the predicted scores generated by our logistic classifier and create a set of distance matrices, one for each canopy in the development set. Each score ranges between 0 (minimum distance) and 1 (maximum distance). We set maximum distance scores for inventor pairs with a different middle name or whose patents were filed more than 40 years apart.

We then apply hierarchical agglomerative clustering (HAC), a method that iteratively clusters inventors within the same canopy based on their respective distance matrix scores. We determine the optimal distance threshold to halt the HAC process (and finalize the set of disambiguated inventors) as the score jointly maximizing precision and recall. In our context, precision quantifies the algorithm's accuracy in grouping inventors correctly. It is calculated as the ratio of "true positives" (correctly clustered inventor pairs) to the total number of true positives and "false negatives" (incorrectly clustered inventor pairs). Recall, on the other hand, measures the algorithm's completeness in identifying all inventors that should be grouped to-

[20]The primary difference between our features and those used by Monath et al. (2021) lies in the method for determining technological overlap between two patents. While Monath et al. (2021) use textual similarity between patent titles, we rely on technological classification codes like the NBER and USPC because titles are not readily available for historical patents.

gether. It is calculated as the ratio of true positives to the sum of true positives and "false negatives" (inventor pairs that should have been clustered together but were not).

Figure B.1 shows the frontier of the precision-recall curve, computed by comparing the disambiguated inventor identifiers generated by the HAC process at several distance threshold increments with the "true" disambiguated status of each pair in the development set. We calculate the optimal distance threshold using the F-beta score, that is, the weighted harmonic mean of precision and recall. We introduce a weight of 0.4 for precision, aiming to privilege the minimization of false positives. We determine an optimal distance threshold of 0.25, associated to a precision score of 0.986 and a recall score of 0.946.

We conclude our procedure by disambiguating all inventors in our dataset, which involves creating the set of inventor canopies and similarity matrices, running the HAC algorithm until the optimal distance threshold of 0.25, and generating a unique identifier for each disambiguated inventor. We count around 2.5 million inventor mentions from historical patents and 5.4 million inventor mentions from modern patents, generating a set of 374 million inventor pairs grouped into canopies. Our final dataset includes 2,352,799 disambiguated inventors associated to 4,653,426 patents.

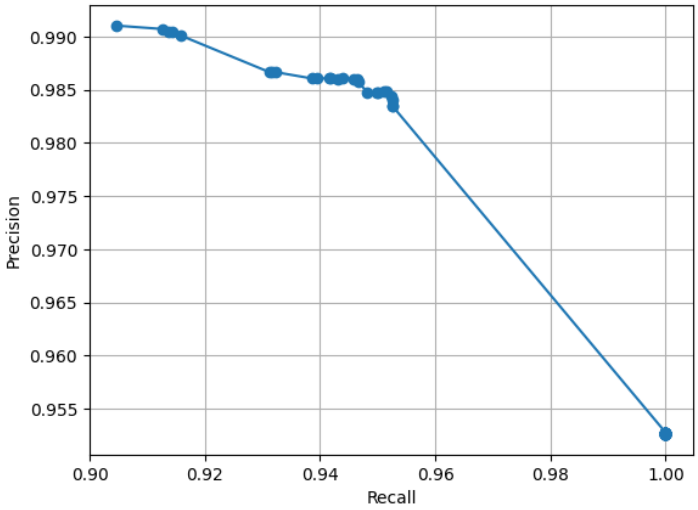
We evaluate the performance of our algorithm through the following steps. First, we focus on inventors from modern patents and assess the overlap between our disambiguated inventor clusters and those of Monath et al. (2021) by calculating the Normalized Mutual Information (NMI) score and the Adjusted Rand Index (ARI). Intuitively, NMI quantifies the similarity between two clustering results of the same dataset, yielding a score from 0 (completely dissimilar clustering) to 1 (perfectly identical clustering). ARI measures the agreement between two clustering results while adjusting for the possibility of random agreement, with values ranging from -1 (completely dissimilar clustering) to 1 (perfectly identical clustering). We obtain an NMI score of 0.995 and an ARI score of 0.960, suggesting that our disambiguation for modern patent inventors is nearly identical to that of Monath et al. (2021).

Second, similar to Akcigit et al. (2022), we search for the top 50 most prolific inventors in our dataset in a list of the most prolific inventors known stored on Wikipedia. A significant challenge for any disambiguation algorithm is distinguishing between homonymous inventors working in the same location and similar technology fields. A low-quality disambiguation process would generate false profiles of top prolific inventors by incorrectly clustering inventors with common names under the same profile. Out of the 50 inventors, we found 45 in the Wikipedia list and confirmed the identities of the remaining five by consulting their biographical profiles on companies' or universities' websites (Table B.1 and Table B.2).

Since Wikipedia tracks inventor careers up to 2024 and includes patents filed at intellectual property offices other than the USPTO, we cannot directly compare the exact number of patents in each inventor's portfolio predicted by our algorithm. Nevertheless, when we focus only on inventors whose entire careers are covered by our dataset, we find minimal differences (Table B.3), both for inventors listed only on modern patents and those with portfolios consisting solely of historical or a mix of historical and modern patents.

Lastly, we extract a random sample of inventor pairs found only in historical patents and asses the quality of the disambiguation by manually inspecting the quality of our disambiguation algorithm's predictions and calculating precision and recall scores.

Figure B2: Precision-recall curve



Notes: Precision and recall scores computed by comparing the "true" disambiguated status of inventor pairs in the development set with disambiguated inventors generated by the HAC process at distance thresholds between 0.01 and 1 with 0.01 increments.

Table B1: Top 50 most prolific inventors manual search results

Name	Total patents	Found on Wikipedia	Found in other sources
Gurtej S. Sandhu	1157	✓	
Leonard Forbes	1074	✓	
Lowell L. Wood, Jr.	1021	✓	
Donald E. Weder	992	✓	
George A. Lyon	908	✓	
Melvin De Groote	882	✓	
Jay S. Walker	853	✓	
Warren M. Farnworth	763	✓	
Edward K. Jung	722	✓	
Roderick A. Hyde	722	✓	
George Spector	710	✓	
Salman Akram	699	✓	
William H. Eby	697	✓	
Austin Gurney	668	✓	
James A. Jorasch	668	✓	
William I. Wood	663	✓	
Michael J. Sullivan	643	✓	
Ahmadreza Rofougaran	630	✓	
Rick A. Hamilton II	626	✓	
Audrey Goddard	622	✓	
Clarence T. Tegreene	615	✓	
Kie Y. Ahn	604	✓	
Paul Godowski	593	✓	
Jeyhan Karaoğuz	567		✓
Mark I. Gardner	515	✓	
Lee D. Whetsel	514	✓	
Edward J. Nowak	511	✓	
Kangguo Cheng	500	✓	
John F. O'Connor	499	✓	
Ravi K. Arimilli	498	✓	
Geoffrey B. Rhoads	492		✓
Nathan P. Myhrvold	490	✓	
Clyde C. Farmer	490	✓	
Anthony J. Baerlocher	489	✓	
Edwin H. Land	485	✓	
Jack A. Mandelman	481	✓	
Frankie F. Roohparvar	478		✓
Mark A. Malamud	473	✓	
Louis H. Morin	469	✓	
Louis L. Hsu	467	✓	
Royce A. Levien	466	✓	
Muriel Y. Ishikawa	445	✓	
Robert W. Lord	442	✓	
David R. Hall	435	✓	
Niall R. Lynman	427		✓
Jeffrey P. Gambino	421	✓	
Shmuel Shaffer	420		✓
James M. Hart	419	✓	
Scott H. Wittkopp	418	✓	
John D. Rinaldo, Jr.	410	✓	

Notes: Results of our search for the top 50 most prolific inventors ranked by our disambiguation algorithm among Wikipedia's top prolific inventors (`https://en.wikipedia.org/wiki/List_of_prolific_inventors` and `https://en.wikipedia.org/wiki/Talk:List_of_prolific_inventors`; last access: September 2024).

Table B2: Biographical details of inventors not found among Wikipedia's prolific inventors

Name	**Biographical extract**	**Source**
Jeyhan Karaoğuz	"Dr. Karaoğuz conducted pioneering research and development in broadband network access, wireless connectivity, and mobile devices in their early days. He is a prolific inventor with 764 US patents awarded to his name."	North Carolina State University, Department of Computer Science (2023)
Geoffrey B. Rhoads	"Mr. Rhoads, the inventor of ZuluTime's technology, started his career at Tektronix, becoming the principal designer of the first commercially available 1 GHz digitizing oscilloscope. He was the chief scientist in working with key customers/applications such as Los Alamos/Livermore in the then nascent commercial efforts toward fiber optics and laser communications systems. In 1994, Mr. Rhoads founded Digimarc Corporation as its Chief Technology Officer and inventor of digital watermarking technology. At Digimarc he led the intellectual property and technology development, directly resulting in over 600 patents awarded to the company to date."	Crunchbase (2024)
Frankie F. Roohparvar	"In his new role, Roohparvar will serve as chief strategy officer. Prior to joining Xitore, he served as VP and senior strategic advisor at HGST (a WD company). He became part of the HGST team as a result of HGST's acquisition of Skyera, an all flash array storage start-up, where he served as the CEO. Prior to joining Skyera, he was the GM and VP of the NAND business unit at Micron Technology. He is a prolific inventor and has over 500 U.S. patents."	StorageNewsletter (2016)
Niall R. Lynman	"Niall R Lynam, a prolific inventor based in Holland, MI, United States, [...] has made significant contributions to the field of vehicular control systems. With an impressive portfolio of 604 patents, Lynam's expertise lies in developing advanced technologies that enhance driver assistance and improve overall vehicle safety."	IDiyas (2024)
Shmuel Shaffer	"Before joining Cisco in 1999, Dr. Shaffer held managerial positions at Siemens ROLM where he headed the Voice over IP (VoIP) development, the U.S. Hardware organization, and the Embedded SW development. Dr. Shaffer received his Ph.D. from Stanford University in the area of Adaptive Self Optimizing Systems. He also holds two Masters Degrees from Stanford in Electrical Engineering and in Statistics. Dr. Shaffer has authored over 250 US patents."	Stanford University, Department of Electrical Engineering (2011)

Notes: The table reports the biographical details and source we used to confirm the identity of the five inventors ranked among the top 50 most prolific inventors by our disambiguation algorithm, but not found among Wikipedia's prolific inventors.

Table B3: Comparison of total inventor patent stocks from our algorithm and Wikipedia's list

Name	**Years of activity**	**Total patents**	**Total patents (Wikipedia)**
Donald E. Weder	1978-2015	992	1000
Melvin De Groote	1924-1966	882	925
George Spector	1974-1998	710	723
Kie Y. Ahn	1970-2013	604	622
Paul Godowski	1994-2010	593	579
Edwin H. Land	1933-1990	485	535
Jack A. Mandelman	1987-2014	481	481
Louis H. Morin	1924-1969	469	503

Notes: The table shows the total number of patents associated to top prolific inventors by our disambiguation algorithm and compares it to the total number of patents found on their profile on Wikipedia's list of top prolific inventors. Since Wikipedia's covers patents filed until 2024, we restrict the sample to inventors whose entire careers are covered by our dataset (1920-2015).

B.3. Inventor Birth Information Search

Patent data lack information on inventors' biographical characteristics such as birth year and birth location, typically reporting only the inventors' full name and city of residence. This complicates linking inventor records generated in adulthood with census information recorded in childhood, which likely report different residential locations. We overcome this challenge by searching birth information for inventors in our sample on *ancestry.com*, enabling us to link inventors active during the second part of the twentieth century to their younger selves in the 1940 census, combining information on inventors' careers and productivity with extensive data on their childhood and parental background.

We begin by selecting all inventors residing in the US and patenting at least once at the USPTO between 1945 and 1990 (985,636 inventors). We select this patenting time window to capture most of the patenting careers of inventors born between the late 1910s and the early 1930s, which are our birth cohorts of interests. To verify whether the 1945-1990 patenting time window is appropriate, we simulate the expected age distribution of inventors active in different patenting time windows.

Specifically, we generate a simulated set of inventors born between 1915 and 1935, ensuring their distribution is proportional to that of US residents observed in the 1970 census.[21] For each inventor in each birth year, we randomly draw the age in which they first patented from the distribution of inventors' first patent age using calculated using data from Kaltenberg et al. (2023).[22] Based on this simulated data, we compare the birth year distribution of inventors we would obtain for a given patenting time window with their expected distribution based on the 1970 census data.

Figure B.2 plots the kernel density estimates obtained by selecting five patenting time windows, all starting in 1945 and ending between 1975 and 1995. The birth year distribution for the 1945-1990 patenting time window closely follows the distribution proportional to the 1970 census, although it slightly underestimates inventors born from 1933 onwards. Shorter patenting windows tend to underestimate the number of inventors born from the mid-1920s onwards.

With evidence that the 1945-1990 patenting window would capture nearly all inventors in our birth cohorts of interest, we search for their birth information on *ancestry.com*. Specifically, for each inventor patenting at least once between 1945 and 1995, we search for their first, middle, and last name, along with their city of residence as listed on the patent. We search for the same inventor more than once only if they resided in different US states throughout their career.

[21] We use three 1 percent samples from the 1970 census available on IPUMS (Ruggles et al. 2021).

[22] Kaltenberg et al.'s 2023 data reports birth year information for the majority of inventors listed on modern USPTO patents (i.e., granted since 1975).

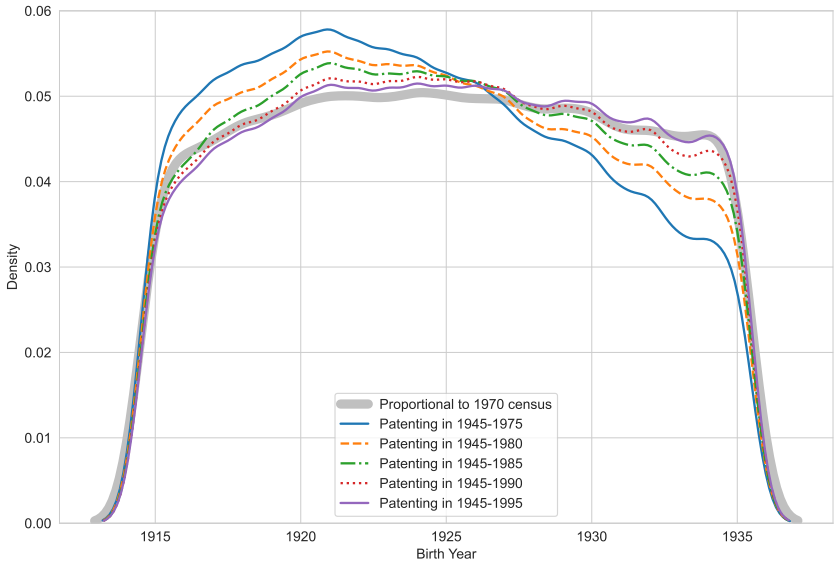
We collect only records reporting the same residence as the inventor, at either the state, county, or city level, depending on the granularity of the geographic information available. Informed by studies on inventors' lifecycle (Jones 2009; Bernstein et al. 2022; Kaltenberg et al. 2023), we keep only records referring to individuals who were between 25 and 65 years old during their patenting career. Additionally, we discard any record reporting a death year before the inventor's last patent.

We retain birth information only for inventors with a unique candidate birth information profile. Using this procedure, we identify birth information for around 56 percent of inventors in our sample (549,768 out of a total of 985,636 inventors). Birth year information is consistently available. Birth location data is available for 60 to 70 percent of inventors born until 1925, a proportion gradually decreasing for younger birth cohorts (from roughly 50 to 35 percent for inventors born between 1930 and 1935, Figure B.3). Such uneven coverage is due to the nature of the records from which we source birth information, more likely to stem from telephone or city directories, typically listing only individuals' birth year and not their birth location. We discuss and address the implications of this uneven birth location coverage on our inventor-census linkage in the next section B.4.

We verify the quality of the inventor birth information data in two steps. First, we manually check a random sample of 300 inventors. For each inventor name, birth year, and birth location combination, we look for information on the individual's residence, occupation, industry, and employer, making sure it corresponds to the associated inventors' residence history, technology field, and employer. We also mark any record which may be associated with homonyms sharing the same birth year and location, as well as the same state of residence. We are able to find birth information for around 73 percent of inventors in the sample and we can confirm correct birth year and location for roughly 95 percent of those records.

Second, we assess how representative the sample of inventors for whom we found birth information is relative to the full inventor dataset we started our search from. Table B.4 reports the mean values and standard deviation for several key inventor characteristics. The differences between the two are minimal across all dimensions we considered, such as the total number of patents filed by inventors during their careers, the location where they filed their first patent, and the technology category where they patented. This suggests that the sample of inventors with birth information is representative of the full inventor sample.

Figure B3: Simulation of inventors' birth year distribution by patenting window



Notes: The plot shows kernel density estimates for inventors' birth year distribution. The gray line represents the distribution proportional to the 1970 US Census. The other lines simulate inventor distributions for five patenting time windows, each progressively larger, starting in 1945 and ending between 1975 and 1995.

Figure B4: Share of records with birth location by birth year

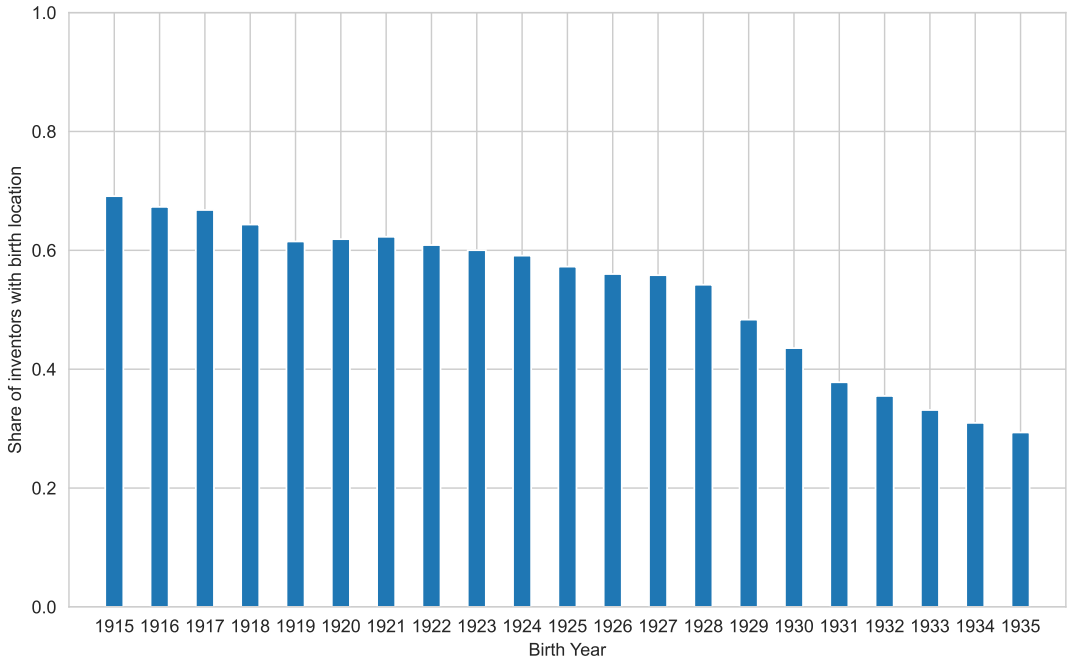


Table B4: Full inventor sample vs. birth information subsample: comparison of means

	Full inventor sample	**Birth information subsample**	**Difference**
Lifetime patents	3.866	3.682	0.184
	(9.24)	(7.98)	
First patent location			
North	0.306	0.293	0.012
	(0.46)	(0.46)	
South	0.202	0.197	0.005
	(0.40)	(0.40)	
Midwest	0.288	0.305	-0.017
	(0.45)	(0.46)	
Mountain west	0.041	0.042	-0.000
	(0.20)	(0.20)	
Pacific	0.160	0.162	-0.002
	(0.37)	(0.37)	
Technology category			
Chemicals	0.234	0.225	0.009
	(0.42)	(0.42)	
Computers	0.118	0.117	0.001
	(0.32)	(0.32)	
Drugs	0.073	0.067	0.006
	(0.26)	(0.25)	
Elec. Eng.	0.228	0.231	-0.003
	(0.42)	(0.42)	
Mech. Eng.	0.339	0.347	-0.008
	(0.47)	(0.48)	
Other	0.393	0.400	-0.007
	(0.49)	(0.49)	
Observations	985,636	549,768	

Notes: The table reports mean and standard deviation (in parentheses) for selected inventor characteristics. The "full inventor sample" includes all inventors who patented at least once in the US between 1945 and 1985. The "birth information subsample" is a subsample of the former, including the inventors for whom we were able to find birth year and birth location information.

B.4. Inventor-Census Linkage

We link inventors born between 1915 and 1935 to the full count 1940 census. We adopt the class of ABE algorithms (Abramitzky et al. 2012, 2014, 2021a), building on their linkage routine to leverage the information provided by inventor and census data effectively. Our main linking features are full names, birth year, and birth location. Stemming from legal documents, inventors' full name strings are highly accurate. In particular, they always report the middle name, if it exists, making it an important feature to identify true matches among multiple candidate links. Our procedure involves the following steps:

1. We parse the the full name strings of both inventors and individuals in the census, removing non-alphabetic characters, correcting misspellings, and accounting for nicknames.
2. We create a set of candidate links by matching inventor and census records on their full first and last name string.
3. We require middle name initials to match, only if they are available for both records. That is, if a middle name is present only for the inventor record, we allow for the possibility that such information may be missing in the census record and retain such pairs.
4. We keep only records with the same birth year or a birth year difference of $\pm$ 1 year.
5. We keep only records with the same birth location (US state or foreign country).
6. We retain only unique inventor-census records pairs. That is, we exclude inventor records with multiple census candidates or multiple inventors linked to the same census candidate.
7. In order to address the progressively lower availability of birth location information for inventors in younger birth cohorts, we add to the sample unique inventor-census pairs with strictly matching middle name initials (or absent middle name for both). Those would have been discarded due to the lack of birth location information for the inventor.

We obtain a sample of 79,019 inventors linked to the census, around 34 percent of our starting sample of inventors born between 1915 and 1935 (235,708 inventors). We refer to this sample as "linked sample (a)".

We produce two alternative linked samples, which we use to test the robustness of our analysis to alternative linkage routines. For the first alternative ("linked sample (b)"), we follow the same procedure up to point 6, keeping in our sample only unique inventors-census pairs sharing the same birth location. This sample includes 43,679 inventors linked to the census (or 19 percent of the starting inventor sample). For the second alternative ("linked sample (c)"), we do not consider birth location information. We follow the same procedure up to point 4 and then retain only unique inventor-census pairs with matching middle name initials (or absent middle name for both). This sample includes 74,286 inventors (or 32 percent of the starting

inventor sample).

We assess the representativeness of each linked inventor sample by comparing some of their key characteristics with our starting inventor sample of inventors born between 1915 and 1935. Table B.5 reports mean and standard deviation values for key inventor characteristics as. The records in linked sample (a) show minimal differences relative the starting inventor sample, displaying a similar average values for birth year and lifetime patents, as well as highly comparable proportions of inventors across first patenting regions and by technology field. Linked sample (c) reports similar values, indicating a high degree of similarity of with both our starting inventor sample and linked sample (a).

The records in linked sample (b) are similar to the starting inventor sample, although they report two key differences. First, they are relatively more represented in the mechanical engineering category and among inventors patenting in the Midwest. Second, they are older, with an average age of 1923.7 compared to 1924.9 for the full starting sample. Such a difference is more evident when comparing the entire birth year distribution of our starting sample and the linked ones (Figure B.4). The distribution of inventors in linked sample (a) and (c) closely tracks that for our starting inventor birth year sample. Linked sample (b) is associated to a more skewed distribution, appearing to be less representative of inventors born from 1928 onwards.

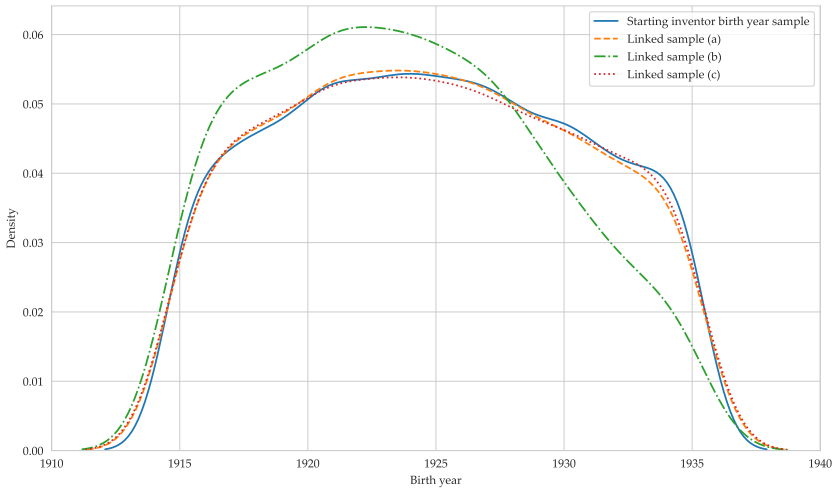
This is explained by the routine generating linked sample (b). By requiring links to report the same birth location, it automatically restricts the sample only to records with available birth location, which are progressively fewer for younger birth cohorts.

Table B5: Full inventor sample vs. inventor-census linked samples: comparison of means

	Starting inventor birth year sample	**Linked sample (a)**	**Difference: Starting - linked (a)**	**Linked sample (b)**	**Difference: Starting - linked (b)**	**Linked sample (c)**	**Difference: Starting – linked (c)**
Birth year	1924.882 (5.76)	1924.795 (5.74)	0.087	1923.737 (5.41)	1.145	1924.833 (5.79)	0.049
Lifetime patents	2.864 (4.84)	2.838 (4.72)	0.026	2.772 (4.64)	0.092	2.836 (4.72)	0.027
First patent location							
North	0.312 (0.46)	0.301 (0.46)	0.011	0.325 (0.47)	-0.013	0.299 (0.46)	0.013
South	0.197 (0.40)	0.188 (0.39)	0.010	0.155 (0.36)	0.042	0.190 (0.39)	0.007
Midwest	0.299 (0.46)	0.322 (0.47)	-0.023	0.357 (0.48)	-0.057	0.317 (0.47)	-0.018
Mountain west	0.037 (0.19)	0.037 (0.19)	-0.000	0.034 (0.18)	0.003	0.037 (0.19)	-0.000
Pacific	0.154 (0.36)	0.152 (0.36)	0.002	0.129 (0.33)	0.025	0.156 (0.36)	-0.002
Technology category							
Chemicals	0.224 (0.42)	0.217 (0.41)	0.007	0.210 (0.41)	0.015	0.219 (0.41)	0.006
Computers	0.092 (0.29)	0.089 (0.28)	0.003	0.078 (0.27)	0.014	0.090 (0.29)	0.002
Drugs	0.049 (0.22)	0.045 (0.21)	0.005	0.041 (0.20)	0.008	0.045 (0.21)	0.004
Elec. Eng.	0.230 (0.42)	0.230 (0.42)	-0.001	0.221 (0.42)	0.008	0.230 (0.42)	-0.000
Mech. Eng.	0.348 (0.48)	0.360 (0.48)	-0.013	0.371 (0.48)	-0.023	0.359 (0.48)	-0.011
Other	0.385 (0.49)	0.387 (0.49)	-0.002	0.396 (0.49)	-0.011	0.385 (0.49)	-0.000
Observations	235,715	79,019		43,679		74,286	

Notes: The table reports mean and standard deviation (in parentheses) for selected inventor characteristics. The "full inventor birth year sample" includes all inventors born between 1915 and 1935, the starting dataset for our linkage procedure. "Linked sample (a)" is our baseline inventor-census linked sample. Linked sample (b) is our first alternative linked sample, obtained by restricting the linkage only to records sharing the same birth location. "Linked sample (c)" is our second alternative linked sample, obtained by restricting the links to records with matching middle name initials (or absent middle names for both), not using birth location information.

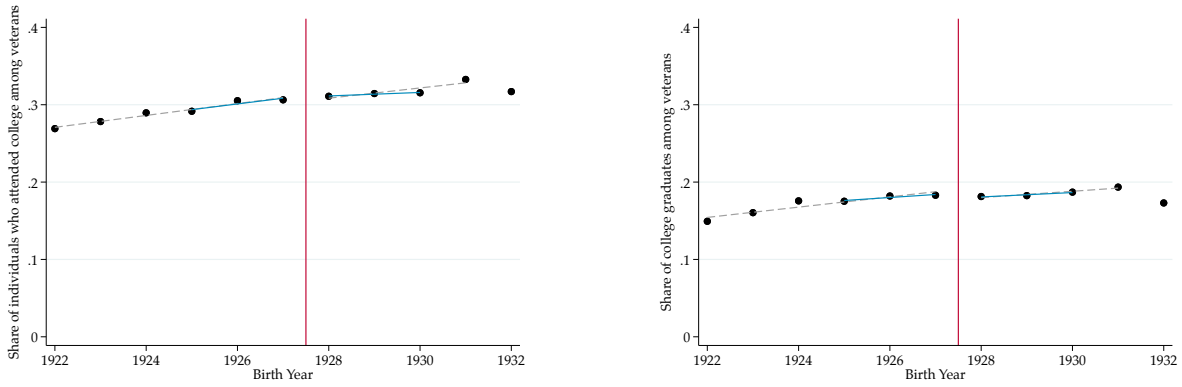
Figure B5: Birth year distribution for inventors in the starting birth year sample and in the three samples linked to the 1940 census



Notes: Kernel density estimates for inventors' birth year distribution. The starting inventor birth year sample includes all inventors with birth information born between 1915 and 1935. Linked sample (a), (b), and (c) are subsets of the starting inventor birth year sample linked to the full count 1940 census.

C. Additional Descriptive Evidence

Figure C1: Share of WWII and Korean War veterans who attended or graduated from college



(a) College attendance

(b) College graduation

Notes: Panel (a) plots the share of WWII and Korean War veterans who attended at least one year of college. Panel (b) shows the share of WWII and Korean War veterans who graduated from college. Both panels are based on an aggregation of three 1 percent samples from the 1970 census. Dashed gray lines show linear fits for cohorts before and after the 1928 cutoff. Blue solid lines show linear fits only for the birth cohorts between 1925 and 1930.

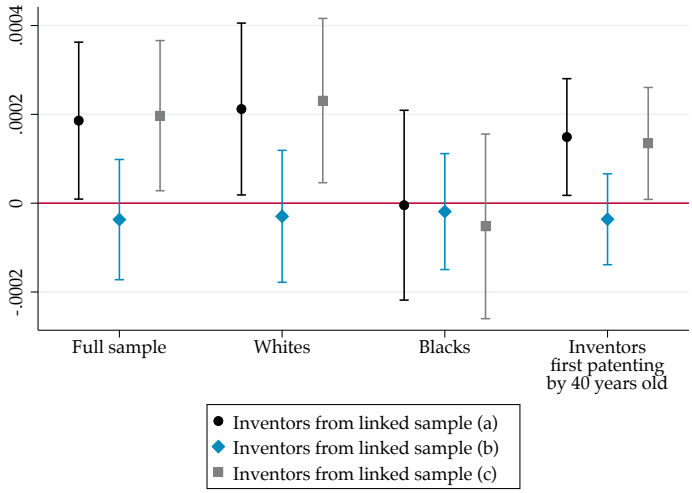
D. Additional Estimates and Robustness Checks

Table D1: Reduced form estimates: individual characteristics

	High school graduate (1950 census) **(1)**	**High school graduate (1970 census)** **(2)**	**Completed 10th grade (1950 census)** **(3)**	**Completed 10th grade (1970 census)** **(4)**	**Resides in urban location (1940 census)** **(5)**	**Distance to closest higher educ. inst. (1940 census)** **(6)**	**Father's annual income (1940 census)** **(7)**
$\mathbb{1}(birth_i < c)$	-0.023***	-0.013***	-0.003	-0.007*	-0.001	0.017	0.497
	(0.002)	(0.005)	(0.002)	(0.004)	(0.001)	(0.027)	(1.483)
Mean dep. var.	0.248	0.583	0.561	0.732	0.0002	0.0009	1058.53
Observations	1,269,905	194,382	1,269,905	194,382	6,063,629	5,759,161	5,929,069
QOB specification	No	Yes	No	Yes	No	No	No

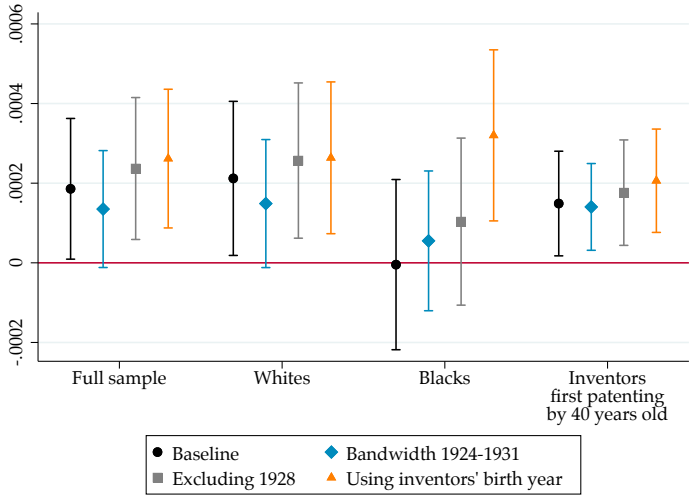
Notes: In columns (1) and (2) the dependent variable is an indicator for individuals who graduated from high school. In columns (3) and (4) the dependent variable is an indicator for individuals who completed 10th grade. In column (5) the dependent variable is an indicator for individuals residing in an urban location. In column (6) the dependent variable is the distance in miles between an individual's residence and the closest higher education institution. In column (7) the dependent variable is the annual income of the individual's father. The bandwidth is 1925-1930 and the birth year cutoff is 1928. All regressions include linear trends before and after the birth year cutoff, birth state dummies, and an indicator for nonwhites. Columns (2) and (4) include also dummies for quarter of birth. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.

Figure D1: Reduced form estimates: probability to become an inventor (alternative linked samples for the dependent variable)



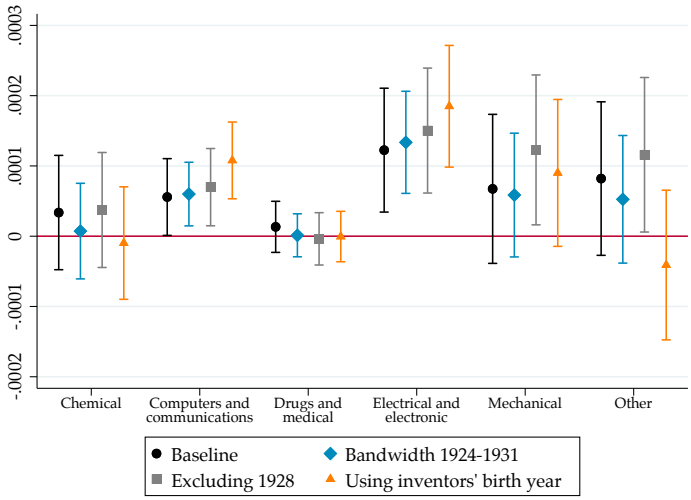
Notes: Each coefficient plots the reduced form estimate and 90 percent confidence interval from equation (2). The dependent variable is an indicator equal to 1 for inventors. Black markers report estimates from regressions where the dependent variable is based on the inventor-census linked sample (a). Blue markers report estimates from regressions where the dependent variable is based on the inventor-census linked sample (b). Gray markers report estimates from regressions where the dependent variable is based on the inventor-census linked sample (c). All regressions include birth state dummies and linear trends before and after the birth year cutoff. Regressions in the "Full sample" and "Inventors first patenting by 40 years old" groups include also an indicator for non-whites. The sample includes all men born in the US between 1925 and 1930 from the full count 1940 census. The birth year cutoff is 1928.

Figure D2: Reduced form estimates: probability to become an inventor (alternative estimates)



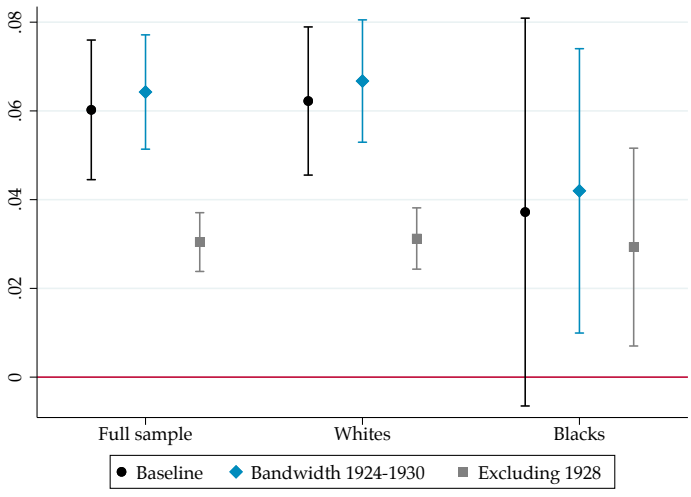
Notes: Each coefficient plots the reduced form estimate and 90 percent confidence interval from equation (2). The dependent variable is an indicator equal to 1 for inventors (based on inventor-census linked sample (a)). Black markers report estimates from baseline regressions with the 1925-1930 birth year bandwidth. Blue markers are based on regression where the birth year bandwidth is extended to 1924-1931. Gray markers report estimates from regressions excluding the 1928 birth year, and considering the birth years 1925, 1926, and 1927 (before the cutoff) and 1929, 1930, and 1931 (after the cutoff). Orange markers report estimates where for inventors we use the birth year identified with our birth information search, instead of the birth year found in the 1940 census. All regressions include birth state dummies and linear trends before and after the birth year cutoff. Regressions in the "Full sample" and "Inventors first patenting by 40 years old" groups include also an indicator for non-whites.

Figure D3: Reduced form estimates: probability to become an inventor by technology category (alternative estimates)



Notes: Each coefficient plots the reduced form estimate and 90 percent confidence interval from equation (2). The dependent variable is an indicator equal to 1 for inventors who patented in a specific technology category at least once (based on inventor-census linked sample (a) and the NBER technology categorization for patents, Hall et al. (2001)). Black markers report estimates from baseline regressions with the 1925-1930 birth year bandwidth. Blue markers are based on regression where the birth year bandwidth is extended to 1924-1931. Gray markers report estimates from regressions excluding the 1928 birth year, and considering the birth years 1925, 1926, and 1927 (before the cutoff) and 1929, 1930, and 1931 (after the cutoff). Orange markers report estimates where for inventors we use the birth year identified with our birth information search, instead of the birth year found in the 1940 census. All regressions include birth state dummies and linear trends before and after the birth year cutoff. Estimates by OLS.

Figure D4: IV estimates: college attendance (alternative estimates)



Notes: Each coefficient plots the IV estimate and 90 percent confidence interval for the effect of veteran status on college attendance. The dependent variable is an indicator equal to 1 for individuals who attended at least one year of college. Black markers report estimates from baseline regressions with the 1925-1930 birth year bandwidth. Blue markers are based on regression where the birth year bandwidth is extended to 1924-1931. Gray markers report estimates from regressions excluding the 1928 birth year, and considering the birth years 1925, 1926, and 1927 (before the cutoff) and 1929, 1930, and 1931 (after the cutoff). All regressions include birth state dummies and linear trends before and after the birth year cutoff. "Full sample" regressions include also an indicator for non-whites.

Table D2: Reduced form estimates (inventors only): probability to patent in a given technology category

	Chemical	**Computers and communications**	**Drugs and medical**	**Electrical and electronic**	**Mechanical**	**Other**
$\mathbb{1}(birth_i < c)$	-0.007	0.016***	-0.001	0.020***	-0.003	-0.012
	(0.007)	(0.005)	(0.003)	(0.007)	(0.008)	(0.008)
Mean dep. var.	0.226	0.101	0.050	0.241	0.339	0.374
Observations	72,328	72,328	72,328	72,328	72,328	72,328

Notes: The sample includes all inventors born between 1925 and 1930. The dependent variable is an indicator equal to 1 for inventors who patented in a specific technology category at least once (based on the NBER technology categorization, Hall et al. (2001)). The birth year cutoff is 1928. All regressions include linear trends before and after the birth year cutoff, birth state dummies. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.

Table D3: Reduced form estimates (inventors only): probability to patent in a given technology sub-category

	Computers and communications				**Electrical and electronic**						
	Communic.	**Computer hardware & software**	**Computer peripherals**	**Information storage**	**Electrical devices**	**Electrical lighting**	**Measuring & testing**	**Nuclear & X-rays**	**Power systems**	**Semiconductor devices**	**Misc. elec.**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$\mathbb{1}(birth_i < c)$	0.008**	0.004*	0.001	0.004*	0.014***	-0.001	0.002	0.002	0.009**	0.004**	0.010***
	(0.004)	(0.003)	(0.001)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
Mean dep. var.	0.061	0.028	0.007	0.017	0.070	0.028	0.060	0.026	0.066	0.012	0.042
Observations	72,328	72,328	72,328	72,328	72,328	72,328	72,328	72,328	72,328	72,328	72,328

Notes: The sample includes all inventors born between 1925 and 1930. The dependent variable is an indicator equal to 1 for inventors who patented in a specific technology category at least once (based on the NBER technology categorization, Hall et al. (2001)). The birth year cutoff is 1928. All regressions include linear trends before and after the birth year cutoff, birth state dummies. Heteroskedasticity-robust standard errors are reported in parentheses. Estimates by OLS. * p<0.1, ** p<0.05, *** p<0.01.