Measuring the Characteristics and Employment Dynamics of U.S. Inventors*

Ufuk Akcigit† Nathan Goldschlag‡

May 7, 2025

Abstract

Innovation is a key driver of long-run economic growth. Studying innovation requires a clear view of the characteristics and behavior of the individuals who create new ideas. A general lack of rich, large-scale data has constrained such analyses. We address this by introducing a new dataset linking patent inventors to survey, census, and administrative microdata at the U.S. Census Bureau. We use this data to provide a first look at the demographic characteristics, employer characteristics, earnings, and employment dynamics of inventors. These linkages, which will be available to researchers with approved access, dramatically increase the scope of what can be learned about inventors and innovative activity.

JEL Codes: O3, O4
Keywords: Inventors, innovation, R&D, firms, dynamism, reallocation

*Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release. DRB Approval Number(s): CBDRB-FY20-CES007-004, CBDRB-FY21-CES007-004, CBDRB-FY22-CES008-008, CBRDB-FY23-CES020-001, CBRDB-FY23-CES020-002, CBDRB-FY24-CES022-006. DMS Project Number 7083300. We thank Joseph Staudt, Shawn Klimek, Emek Basker, anonymous referees, participants at the 2022 Research Data Center Annual Conference and USPTO Chief Economist Seminar for helpful comments and suggestions.

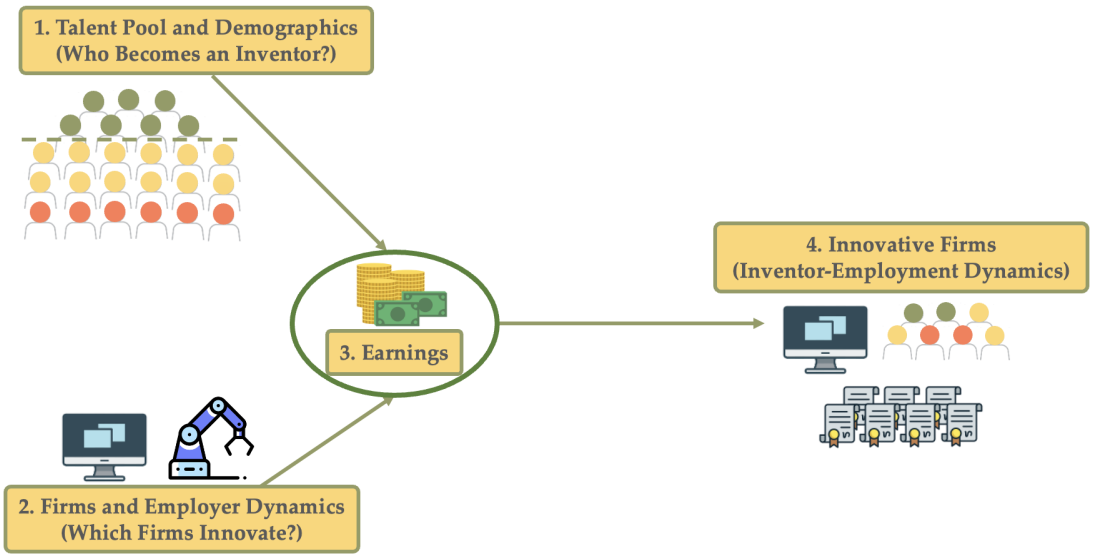
†University of Chicago, NBER, CEPR

‡U.S. Census Bureau

1 Introduction

New ideas drive economic growth, and yet, ideas do not fall like manna from heaven–they are conceived by individuals. Firms that wish to innovate must compete for scarce inventive talent. History is packed with stories of inventors walking away from their employers over pay disputes. Nikola Tesla's departure from Edison's company might be the most iconic. Frustrated by various conflicts, including over compensation, Tesla famously wrote in his diary on December 7, 1884: "Goodbye to the Edison Machine Works." This moment captures a powerful truth: financial reward plays an important role in where inventors choose to work, shaping the entire innovation landscape. Each year, a new wave of inventors enters the market, with a diverse set of firms competing to hire them. The forces of supply and demand, guided by the invisible hand, set the "price" of inventor talent. But how exactly is this talent allocated, and what do inventors actually earn? In this paper, we introduce a novel dataset linking U.S. inventors to Census Bureau survey and administrative data on individuals and their employment histories. We believe this new data will fuel further research into the complexity of the innovation ecosystem.

FIGURE 1: INVENTOR ECOSYSTEM



We utilize this novel dataset to assess the innovative capacity of the U.S. economy by examining four building blocks. Our analysis starts with inventors and their employers, as shown in Figure 1. Innovation is not evenly distributed among individuals or firms; only a subset of individuals become inventors, and only certain firms participate in

innovative activities. Accordingly, our analysis focuses on who becomes an inventor (part 1) and which firms are most likely to innovate (part 2). Additionally, inventors and firms are connected through a price mechanism–a dynamic that has remained largely unexplored due to the absence of large-scale administrative data (part 3). The main contribution of this paper is the creation of a novel, matched employer-employee dataset for inventors, enabling us to document the earnings that facilitate the alignment between inventors and firms. Crucially, once this match is established, inventors often move between employers, leading to a reallocation of innovative talent across firms (part 4). To thoroughly examine the innovative capacity of the U.S. economy, we structure our analysis around four key dimensions: (i) inventor demographics, (ii) firm characteristics, (iii) inventor earnings, and (iv) employment dynamics of inventors. Together, these elements offer a comprehensive view of the human capital driving innovation and how this talent is distributed and reallocated across the economy.

This study presents four key findings about the dynamics of inventors and innovation, motivated by the organizing framework of Figure 1. First, it explores the shifting demographics of inventors, noting the continued underrepresentation of women and minorities, while highlighting significant changes, such as a two-thirds increase in the share of Asian inventors from 2000 to 2016. It also emphasizes the impact of immigration policies on overall innovative activity. Second, the study examines the types of firms that innovate, finding that inventors are increasingly concentrated in older, larger firms, though those at younger firms tend to produce more impactful innovations. This suggests that understanding which firms employ inventors can inform targeted innovation policies. Third, the distribution of inventor earnings is highly skewed and closely linked to productivity and collaboration, with important implications for incentives to innovate and tax policy. Lastly, the study analyzes employment dynamics, noting a decline in inventors starting new firms or changing jobs and an increasing geographical concentration of inventors, which has implications for place-based innovation policies. Together, these findings, summarized below, offer valuable insights into the allocation of inventive talent and the broader innovation ecosystem.

Access to high-quality data has become critically important as researchers and policymakers place greater emphasis on industrial policy and policy evaluation. The empirical evidence presented in this paper highlights the importance of the utilization of inventive talent and financial resources to foster technological progress and drive economic growth. These findings are informative to policymakers aiming to optimize resource allocation in ways that enhance innovation capacity. In the concluding section of this paper, we delve deeper into innovation policy discussions, offering insights to support

effective decision-making in this dynamic landscape.

The following summarizes the basic empirical patterns in the data.

Inventor Demographics

1. **Females are under represented among inventors, especially on a citation weighted basis.** Females account for less than 12% of inventors, a share that is rising and tends to be higher among young inventors. The share of citations accounted for by female inventors lags behind their share of the inventor population. (See Figure 2)

2. **Many inventors are foreign born.** Over 30% of inventors are foreign born and we see a rise and fall of foreign born share among young inventors. China and India account for an increasing share of foreign born inventors, rising from 25% to 40% by 2016. (See Figure 3, 4)

3. **Inventors are getting older.** The average age of inventors rises from 43 to 46 over this period. The share of young inventors falls from nearly 20% in 2000 to under 14% in 2011, but begins to rise thereafter. (See Figure 5)

4. **African Americans are significantly under represented among inventors, Asians inventors are increasingly common.** We find under representation of most minorities, especially for African Americans, who account for less than two percent of inventors. Asians, on the other hand, rose from 13% to 22% of inventors. (See Table 4)

5. **Representation of different demographic groups varies significantly by sector.** Female inventors have greater representation in the Health Care and Social Assistance and Education sectors. Foreign born inventors are more common in the Information, Education, and Professional/Scientific Services sectors. The share of young inventors is highest in the Information sector. (See Table 5)

Employer Characteristics

6. **Inventors work at older, larger firms.** Over 68% of inventors work in firms over 20 years old. Almost 63% of inventors work in large firms with at least 1000 employees. Akcigit and Goldschlag (2023) show that the share of inventors at incumbent firms rose from 49% to 58% between 2000 and 2016. This pattern appears to go back to the 1970s. The share of inventors on patent grants with the largest assignees between 1980 and 2018 rose from about 34% to 47% percent. (See Table 6, 7 and Figure 8)

7. **Inventors at young firms produce more impactful patents.** Inventors at young, medium size firms tend to have the highest impact patents. Inventors at older, smaller firms tend to have the lowest impact patents. (Figure 9)

8. **Inventors, especially the most productive, are less likely to work at young firms.** The share of inventors working at firms ≤ 5 years old fell by almost half over this period from 15% to under 8%. This share fell the most among superstar inventors, those with the most impactful patents. (See Figure 7, 10)

Inventor Earnings

9. **Inventor earnings is highly skewed, especially for superstar inventors.** About 63% of all inventors, and 88% of superstar inventors, are in the top 10% of the earnings distribution. Almost 8% of all and 19% of superstar inventors are in the top 1%. (See Table 8)

10. **Inventor earnings is closely tied to inventive productivity.** Inventors in the top 10% of the inventor earnings distribution have average citations per grant about 0.5 log points higher than inventors in the bottom 10%. Inventors with more collaborators also have higher earnings. (See Figure 12, 13)

Inventor Employment Dynamics

11. **Inventors are less likely to switch jobs over time.** The hire and separation rates for inventors fell from about 6% and 7% respectively in 2000 to 4% in 2016. (See Figure 14)

12. **Inventors, especially superstar inventors, are less likely to start a firm over time.** The probability a superstar inventor becomes an entrepreneur fell by 57%. The quality of inventor founded firms, as measured by post-entry growth, is stable over time. (See Figure 15, 16)

13. **Inventors are increasingly geographically concentrated and less likely to change employment across state lines.** The share of inventors working in the 20 largest counties by inventor count rose from 39% to over 47% between 2000 and 2016. The share of inventors switching employment across state lines fell from a peak of 4.6% in 2006 to 2.6% in 2016. (Figure 17, 18)

The paper proceeds as follows. Section 2 describes how this work relates to the existing literature. Section 3 describes the matching procedures, quality of the resulting links, and the data used to describe the basic facts of U.S. inventors. Section 4 describes

presents in greater detail the demographic characteristics, employer characteristics, earnings, and employment dynamics of inventors. Section 5 concludes.

2 Literature

A large literature characterizes innovation using measures of the R&D and patenting activity of firms (Griliches, 1992, 1998; Hall, 1996). While inventors themselves have long been of interest (Winston, 1937; Schmookler, 1957), only recently have researchers been able to assemble large administrative datasets capable of characterizing the population of inventors. Recent work using such datasets has shed light on a number of important topics including inventor's contribution to long run growth, the relationship between innovation and income inequality, the relationship between talent, human capital, and innovation, the individual returns to innovative activity, the role of team-specific human capital, the processes by which individuals become inventors, the role of parental resources, contributions of IQ, and the role of childhood environments.[1] In this paper, we increase the set of research questions that can be explored by linking inventors to rich survey, census, and administrative data at the U.S. Census Bureau. These data will be available to approved projects via the Federal Statistical Research Data Centers.[2]

Our work also relates to a number of studies that document the relationship between capital or labor misallocation and productivity losses (Hopenhayn and Rogerson, 1993; Hsieh and Klenow, 2009). Relatedly, there can be misallocation of talent to innovation due to socioeconomic background, financial frictions, and R&D and education policies (Bell et al., 2019; Hsieh et al., 2019; Akcigit et al., 2024; Celik, 2023; Aghion et al., 2023). Our stylized facts relate to this work by documenting the allocation of inventive talent in the economy. We find significant patterns of under representation of minority groups, which could suggest a misallocation of talent in the supply of inventors. We also find that inventors at large, young firms tend to have the most impactful innovations while the share of inventors at young firms or founding startups has declined, which could result in the misallocation of innovative resources across firms. Akcigit and Goldschlag (2023) focus on issues of allocation more explicitly, showing that not only are inventors at young firms producing more innovative output, but when inventors are hired by incumbent firms they earn more and innovate less.

[1] For example, see Akcigit et al. (2017), Akcigit et al. (2024), Toivanen and Väänänen (2012), Aghion et al. (2018), Kline et al. (2019), Jaravel et al. (2018), Aghion et al. (2017), Celik (2023), Bell et al. (2019), Aghion et al. (2023).

[2] See `https://www.census.gov/about/adrm/fsrdc.html`.

As noted previously, earnings plays the crucial role of coordinating the actions of inventors and innovative firms. Understanding the distribution of inventor earnings is important for a number of reasons. Earnings incentivizes inventive effort, leading to earnings premiums that scale with the quality of inventive output (Toivanen and Väänänen, 2012). Firms also share innovation rents with difficult to replace inventors to increase retention (Kline et al., 2019). Moreover, where inventors sit on the earnings distribution may have implications for how tax policies affect innovative activity (Akcigit et al., 2016). Our inventor employment histories are uniquely suited to studying the dynamics of inventor earnings. We show that inventor earnings are highly skewed and closely tied to measures of inventive productivity.

This paper also relates to the firm dynamics literature. In the past several decades the U.S. economy has exhibited trend declines in startup rates and employment flows, rising concentration, and falling productivity growth (Decker et al., 2014, 2020; Akcigit and Ates, 2021; Andrews et al., 2016). Slower knowledge diffusion and the strategic behavior of market leaders plays an important role in explaining those trends (Akcigit and Ates, 2023). In this paper we show declining entrepreneurship rates, declining employment flows, and rising concentration of inventors in older, larger firms. The flow of inventors between firms not only reallocates talent to different projects and tasks, but also transfers tacit knowledge between firms and teams of inventors. As Oppenheimer noted, the best way to send information is to wrap it up in a person.[3] Furthermore, incumbent firms can limit knowledge diffusion through strategic hiring and hoarding of inventors. Therefore, the employment dynamics we report here are of first-order importance to understanding the innovative capacity of the economy.

Finally, a growing body of literature develops models that emphasize the critical role of knowledge diffusion through human interactions in driving economic growth (Lucas Jr and Moll, 2014; Perla and Tonetti, 2014; Akcigit et al., 2018). These studies argue that individuals' productivity–and consequently their earnings–depend on the ideas they acquire through interactions with others. This "learning-by-meeting" mechanism suggests that individuals accumulate knowledge by engaging with those who possess different ideas or expertise, ultimately leading to a more productive workforce. In this study, we provide the first direct evidence of this effect among U.S. inventors, showing that inventor earnings scales with the number of co-inventors.

[3]See the 1948 Time Magazine article "The Eternal Apprentice" (Anonymous, 1948).

3 Matching Inventors

The Census Bureau assigns disambiguated, anonymized, person-level identifiers called Protected Identification Keys (PIKs) to survey and administrative data that include individuals. Matching patent inventors to PIKs allows us to link inventors to other Census Bureau microdata such as the Decennial Census, American Community Survey (ACS), the Longitudinal Employer Household Dynamics (LEHD), and W2 records. This paper introduces an inventor record-to-PIK database covering US-based inventors associated with patents granted since 2000.[4] Patent inventor records are unique patent number-inventor combinations found in the United States Patent and Trademark Office (USPTO) data. Inventors on patent grants that have more than one inventor are differentiated by a unique counter, the inventor sequence number, which captures the order an inventor appears in the patent document.

The bulk of our inventor record-PIK database is made up of matches drawn from the Business Dynamics Statistics of Patenting Firms (BDSPF) triangulation matching methodology.[5] Inventor record-PIK matches are generated by the BDSPF triangulation method when a "fully traingulated" match is made. This triangulation process simultaneously resolves ambiguity in the person matches and firm matches, which often contain multiple false positives. A fully triangulated match occurs when, for a given patent, an inventor record-PIK match and an assignee-firm match are corroborated by a PIK-firm employment spell in the LEHD data. These represent the highest quality assignee-firm and inventor record-PIK matches in our data because matches for the inventor and assignee are validated by an employee-employer linkage in the LEHD data.

We also incorporate unique matches output by the Person Validation System (PVS). The Census Bureau's PVS is used to assign an anonymous, person-level identifier to individuals.[6] Typically, the PVS matches an input file with personally identifiable information (PII) such as name, date of birth, and in some cases SSN, to a confidential reference file. The reference file maps individual PII to a unique PIK. One input of the BDSPF triangulation match are PVS inventor matches that rely on the relatively limited information available within patent documents—the person name, city, and state. The limited PII available for inventors produces relatively ambiguous matches. Many inventor records match to multiple PIKs. Numerous false positives from the PVS match is a

[4]The set of years available for the match is limited by the availability of comprehensive LEHD data. Foreign inventors are not at risk of matching Census Bureau data.

[5]See Appendix A.1 for a brief description of the match, which is described in greater detail in Graham et al. (2018) and Dreisigmeyer et al. (2018).

[6]See Wagner and Lane (2014) for details.

feature, not a bug, for the triangulation match. Additional assignee-firm and person-firm (job) links are used to disambiguate the relatively messy matches. In some cases, however, either when the geography is relatively small or the inventor name is relatively unique, the PVS will assign a unique PIK to an inventor record. When available, we incorporate these unique matches into our inventor record-PIK database.

Finally, we augment our inventor matches using disambiguated inventor identifiers from the PatentsView (PV) database. The PV database contains inventor identifiers intended to link inventor records across patent grants that represent the same individual (Monath et al., 2021). Those identifiers are constructed using inventor information such as name and geography along with co-invention networks. Inventor records with the same (or similar) name that appear across different patent documents with the same (or similar) team of inventors will be linked together. Importantly, our triangulated and PVS matches described above assign PIKs to inventor records, not disambiguaged groupings of inventor records that belong to the same, disambiguated person. Given that, we expect there to be inventor records associated with the same person, but that we were unable to assign a PIK (perhaps because we were unable to match the assignee to a firm in the Business Register). In these cases we can utilize the PV disambiguation to extend the reach of our inventor record-PIK matches. Whenever a PIK uniquely matches to a PV inventor identifier (InvID), and that InvID links to inventor records we were unable to match, we assign the PIK to those additional inventor records.[7]

Ultimately, of the 4.9 million US-based inventor records at risk of a match we are able to assign a PIK to 4.5 million (93 percent). Table 1 shows the distribution of match types.[8] The vast majority (67 percent) of our inventor record matches are derived from BDSPF triangulation matches. About a quarter of our inventor record matches come from unique PVS matches and another 8 percent from extensions using the PV inventor identifiers. The composition of match types remains stable over time.[9]

There are some systematic differences in the characteristics of patents for which we are able to match an inventor record and patents we do not. Table 2 shows several patent-level metrics associated with matched and unmatched inventor records. On average, matched inventor records are associated with patents with more than twice as many citations made and nearly twice as many citations received. Matched inventor records

[7]Note that we do not match additional PIKs using the PV information. The PV identifiers simply allow us to cover additional inventor records with our already matched set of PIKs. We only utilize PV identifiers to which our PIK matches map uniquely. Note also that a similar process could be used to extend our matches prior to 2000, finding patents granted prior to 2000 associated with an InvIDs with a matched PIK.

[8]The match rate itself is quite stable over time, see Figure A2 in Appendix A.2.

[9]See Figure A3 in Appendix A.1.

Table 1: Inventor Record Match Type

Match Type	Percent of Matches
Triangulation, EIN & firmid (A1)	62
Triangulation, firmid (A2)	4
Triangulation, EIN (A3)	1
PVS Inventor-NUMIDENT Match	25
PatentsView Extension	8

Source: Inventor Record-PIK Match, LEHD, PatentsView
Notes: A1, A2, and A3 are triangulation match types in which the PIK is assigned with both EIN and FIRMID matching between the job and BR-matched firm, only FIRMID match, and only EIN match respectively. PatentsView is the share of inventor records matched by using the disambiguated inventor identifiers in the PatentsView database to extend PIK matches to unmatched inventor records. PVS is the share of inventor records matched using unique PVS matches.

are also associated with more independent claims and slightly larger teams (number of inventors). Differences in team size are likely driven by the fact that patents with more inventors are more likely to receive a match for at least one inventor record.

Table 2: Matched and Unmatched Inventor Record Characteristics

Variable	Unmatched	Matched
Cites Made	12.5	30.83
	(30.93)	(75.71)
Cites Received	3.552	6.638
	(10.47)	(19.34)
Independent Claim Count	2.557	3.103
	(2.095)	(2.284)
Team Size	2.367	2.758
	(1.753)	(1.938)

Source: Inventor Record-PIK Match, PatentsView
Notes: Shown are mean patent-level characteristics of unmatched and matched inventor records. Standard deviations in parentheses. Cites Made is the count of citations made. Cites Received, 5 Year Window is the count of citations received within the first five years after the patent grant. Independent Claim Count is the count of independent patent claims as developed by Marco et al. (2019). Team Size is the count of inventors on the patent.

Whenever our BDSPF triangulation and PVS matching assigns a single PIK to mul-

tiple inventor records we necessarily disambiguate inventors in the patent data. This provides us with the opportunity to compare two disambuguated identifiers: the PIKs assigned by BDSPF triangualtion and PVS matching and the PV InvIDs. These two disambiguation methods are quite different but yield very similar groupings of inventor records. Since the groupings are not identical, there is a many-to-many mapping of InvIDs to PIKs. We are able to assess the ambiguity of the groupings in "both directions". That is, we can measure how often InvIDs map to multiple PIKs and how often PIKs map to multiple InvIDs.

The left panel of Table 3 shows the share of InvIDs that map to 1, 2, 3, 4, and 5 or more PIKs. As shown in the second column, about 94 percent of InvIDs map to a single PIK and almost 5 percent map to two PIKs. Less than 1 percent of InvIDs map to more than two PIKs. Among the InvIDs that map to multiple PIKs, we can evaluate how concentrated that mapping is by measuring how many inventor records are associated with the modal InvID-PIK pair. For example, if an InvID maps to two PIKs, one of those InvID-PIK pairs could account of over 90 percent of inventor records for that InvID or they could be split roughly equally. In the third column of Table 3 we see that among the InvIDs that map to two PIKs, one of those PIKs accounts for about 71 percent of the inventor records. Even among the InvIDs that map to five or more PIKs the inventor records tend to be concentrated in one of the PIKs—the PIK with the most inventor records accounts for over half of inventor records.

Table 3: Agreement between PatentsView Inventor ID and PIKs

PIK Match Count	InvID Share	Modal PIK Share	InvID Match Count	PIK Share	Modal InvID Share
1	94.47	100	1	96.63	100
2	4.962	71.04	2	2.725	72.04
3	.5	62.32	3	.4005	62.14
4	.05844	58	4	.1202	54.39
5+	.01181	55.55	5+	.1248	45.58

Source: Inventor Record-PIK Match, PatentsView
Notes: The InvID Share column shows the share of PatentsView inventor IDs that map to some number of PIKs given in the PIK Match Count column. PIK Share shows the share of PIKs that map to some number of InvIDs given in the InvID Match Count column. The Modal PIK Share column shows the share of inventor records accounted for by the most frequent PIK. Similarly, the Modal InvID Share column shows the share of inventor records associated with the most frequent InvID.

The right panel of Table 3 describes ambiguity in the "other direction". We find slightly less ambiguity in the mapping of PIKs to InvIDs. Nearly 97 percent of PIKs

map to a single InvID and about 2.7 percent map to two InvIDs. Again, fewer than one percent of PIKs map to more than two InvIDs. We see a similar pattern of concentration in the inventor records. The InvID match with the most inventor records accounts for 72 percent of records for PIKs that map to two InvIDs and about 46 percent among PIKs that map to 5 or more InvIDs.

Our inventor record-PIK database provides a high quality bridge between inventor records in the patent data to unique person identifiers in the Census Bureau data.[10] With those links we are able to combine survey, census, and administrative data on individual inventors. The overwhelming majority of our PIK matches come from high quality BDSPF triangulation matches. Moreover, we observe substantial overlap between the PatentsView InvIDs and our PIK matches, a positive signal about both the quality of the PatentsView disambiguation algorithm and the PIK matches. The inventor record-PIK database will be made available to researchers with approved access.

3.1 Inventor Jobs Panel

In order to measure the demographics and employment dynamics of inventors we link our inventor-PIK matches to the LEHD data to create an inventor jobs panel. This panel, which we term the Inventor Employment History panel, contains the full employment history of all persons linked to at least one patent grant in our sample period. The LEHD data is derived from quarterly state unemployment insurance (UI) records covering about 95 percent of private sector employment as well as state and local government employment. Not all states are available in the LEHD data every period. We restrict to 45 states with employment data starting in 2004 or earlier. Our panel includes only dominant, beginning-of-quarter (BoQ) jobs from 2000Q1 to 2016Q3.[11] After 2000, the addition of states to the frame introduces only minimal bias in national of job-to-job statistics (Henderson and Hyatt, 2012) and have very little impact on measures of inequality (Abowd et al., 2018). Our earnings measure, in addition to direct compensation, for most states includes bonuses, stock options, severance pay, tips, and other gratuities.[12]

To each quarterly job observation we attach measures of inventive activity by combining our inventor record-PIK links with patent information from PatentsView. Though our jobs panel ends in 2016, we use patent grants through 2019, extending the range of

[10] See Appendix A.2 for additional discussion of match quality.

[11] A beginning-of-quarter job is defined as a job for which the individual has positive earnings in both t and $t-1$. A dominant beginning-of-quarter job is the highest earning beginning-of-quarter job for the individual within a given quarter.

[12] See the BLS Quarterly Census of Employment and Wages documentation for details `https://www.bls.gov/cew/overview.htm`.

our citations received measures. Each quarter we track the number of patent and citation weighted applications and grants associated with the inventor. Our preferred patent citation measure counts external citations received, or citations associated with a different assignee, in the five years after a patent is granted. Separately, in order to identify more incremental innovations, we track the number of self citations received in the same five year window. Finally, we use the independent claim counts developed by Marco et al. (2019) to measure the scope of each patent.

It is worth noting what our data does and does not capture. As described by Griliches (1998), patent statistics can be used to make inferences about innovative activity but care must be taken in their interpretation. For instance, many innovations are not patented (Cohen et al., 2000) and many patents are not valuable (Bessen, 2008). Consistent with this, though many patenting firms introduce new products, many product introductions occur in the absence of a patent (Argente et al., 2020). Despite this, after limiting to high quality patents, patent measures are associated with growth and reallocation and accounts for some of the fluctuations in total factor productivity (Kogan et al., 2017). With these issues in mind, our inventor database captures a set of individuals that are very likely engaged in innovative activity, even if they do not represent the universe of individuals engaged in innovative activity.

4 Results

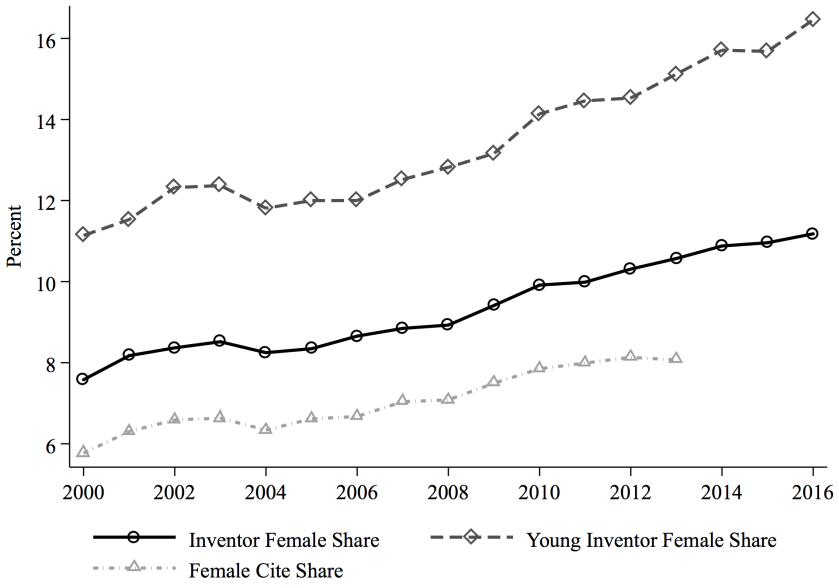
4.1 Inventor Demographics

Consistent, population-level measures of the demographic characteristics of inventors is key to understanding patterns of representation of different groups. Policies aimed at addressing gaps in representation rely on such measurements to gauge the scope and efficacy of policy interventions. In this section, we present a series of demographic measures of the inventor population including age, sex, race, ethnicity, and foreign born status. Future work, enabled by the linkages used in these analyses, can evaluate the link between inventor characteristics and outcomes such as inventive productivity, earnings, and entrepreneurship.

Fact 1: Females are under represented among inventors, especially on a citation weighted basis.
Consistent with previous work, we find under representation of females among inven-

tors. Figure 2 shows the female share of grant active inventors between 2000 and 2016.[13] The share of female inventors rose from about 8 percent in 2000 to over 11 percent in 2016. This is less than the 13.1 percent found by Bell et al. (2019), similar to the 10 percent reported by USPTO (2019), and higher than the 7.9 percent found by Miranda and Zolas (2017).[14] Among young inventors, those less than 36 years old, the female share rose from about 11 percent to over 16 percent—still far below parity. We also find that female inventors tend to be associated with lower-citation patents. Females account for about 10 percent of inventors but only about 7 percent of citations received.[15]

Figure 2: Inventor Female Share



Source: Inventor Employment History
Notes: Female share is the share of grant positive inventors that are female. Young inventor female share is the share of inventors ≤ 35 that are female. Female cite share is the share of cites associated with female inventors. Shares are averaged across quarters within each year.

The disparity in gender representation and citation rates suggest that the innovation

[13]Grant active inventors are those that receive at least one patent grant in the period.

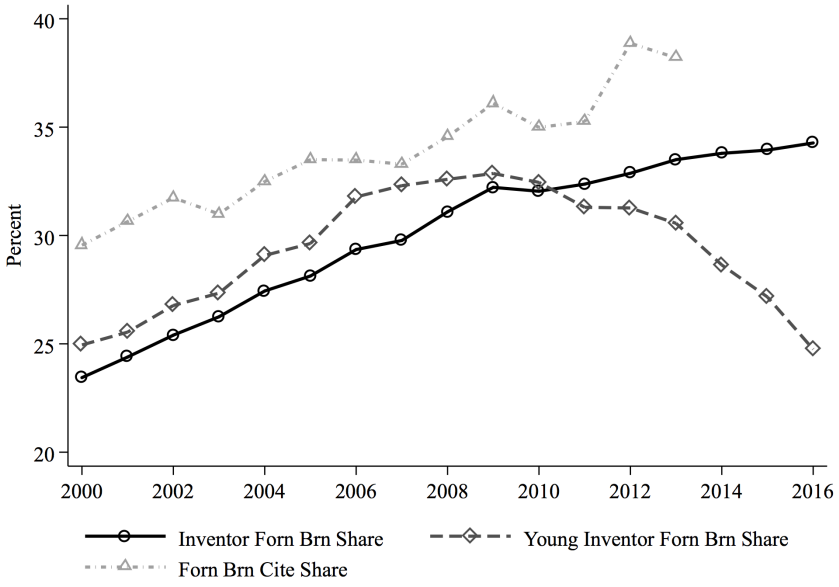
[14]Miranda and Zolas (2017), who focus on the role of household and non-employer inventor activity, utilize PVS matched inventors without the benefit of the Graham et al. (2018) triangulation match. Moreover, the demographic shares reported by Miranda and Zolas (2017) are also not directly comparable to those reported here because they reflect patent-level averaged characteristics and not percentages of inventors.

[15]The citation series is censored at 2013 because we require five years of subsequent citation data. Lower shares of citations, relative to share of inventors, could be related to technology field and industry composition.

system in the U.S. is not fully utilizing the potential of a significant portion of its talent pool. Innovation thrives on diverse perspectives, which means that a more balanced gender representation could lead to more resilient and varied technological progress.

Fact 2: Many inventors are foreign born.
Our data also allow us to identify whether an inventor was born outside of the United States. Figure 3 shows the share of grant active inventors each year that are foreign born, which rises from about 24 percent to almost 35 percent. In contrast to female inventors, foreign born inventors tend to be on higher-citation patents. The foreign born share of citations received rises from about 30 percent in 2000 to nearly 40 percent in 2013. We also find a decline in share of young inventors that are foreign born, initially rising from 25 percent to over 30 percent, but then declining back to 25 percent after 2009.

FIGURE 3: INVENTOR FOREIGN BORN SHARE

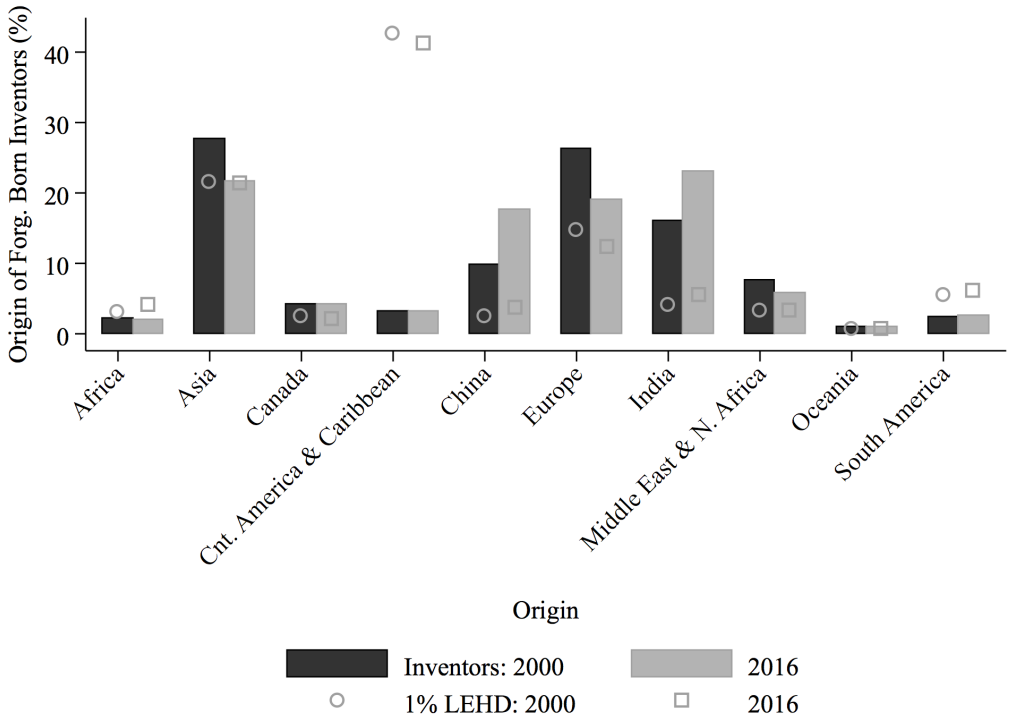


Source: Inventor Employment History
Notes: Foreign born share is the share of grant positive inventors that are foreign born. Young inventor foreign born share is the share of inventors ≤ 35 that are foreign born. foreign born cite share is the share of cites associated with foreign born inventors.

In addition to foreign born status, we also observe the origin region, and in some cases country, of foreign born inventors. In Figure 4 we show how the share of foreign born inventors by origin changed from 2000 and 2016. For purposes of comparison, we also show the share of foreign born workers (not just inventors) by origin region

using a 1% sample of the LEHD. We see that the share of foreign born inventors from China and India rose significantly over this period. Combined, China and India went from just over 25 percent of all foreign born inventors in 2000 to over 40 percent in 2016. The increase in foreign born inventors from China and India is more dramatic than what we see for foreign born inventors in general. Despite the count of inventors from other Asian countries and Europe rising over this period, both fall as a share due to the significant increase in flows from China and India. Finally, comparing overall population and inventor shares, we see that one of the largest gaps in representation among inventors is the Central American and Caribbean group. Over 40 percent of foreign born workers originate from Central America and the Caribbean while less than 4 percent of inventors originate from there.

Figure 4: Origin of Foreign Born Inventors



Source: Inventor Employment History
Notes: Share of grant positive foreign born inventors by origin region/country for 2000 and 2016. Country and region groups are recoded combinations of the Place of Birth codes outlined in Vilhuber (2018). Asia, for example, is all Asian countries excluding China and India. Specifically, Asian includes [1,2,C,D,M,N,P]. Europe includes [F,J,U,X,Y,Z]. Central America & Caribbean includes [4,5,B,G,H,I,O,Q,S,W]. South America includes [6,T]. Middle East & Africa includes [3,V]. The 1% LEHD sample is the employment history of a 1% sample of workers in the full LEHD data.

The striking rise in inventors from China and India could be linked to both countries' increasing focus on higher education, particularly in engineering and technology, coupled with U.S. immigration and employment policies that favor highly skilled workers in these fields.

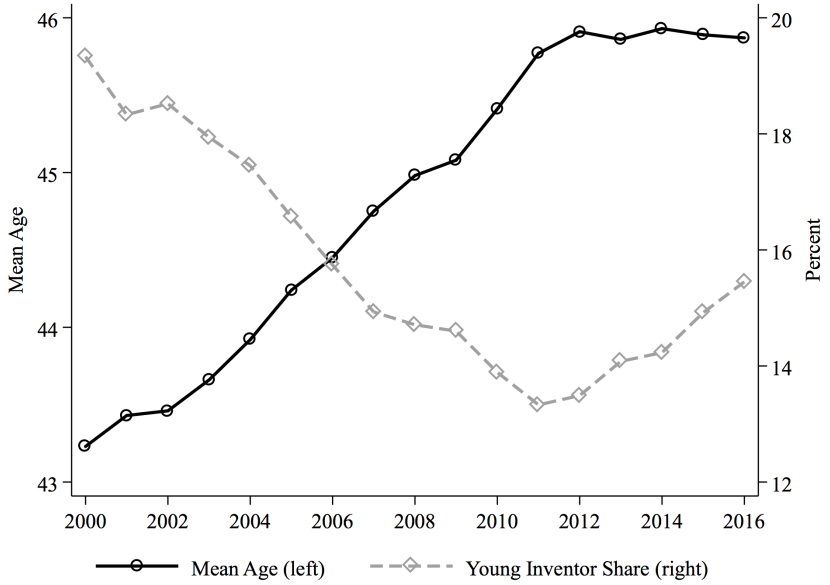
Fact 3: Inventors are getting older.
The age of the U.S. labor force has risen significantly over the last few decades. The share of workers age 55 or older rose from 13.1 percent in 2000 to 23.6 percent in 2020.[16] This trend may have important affects for innovative activity given the life cycle patterns of innovation and entrepreneurship (Jones, 2010; Azoulay et al., 2020). Inventors show similar patterns of aging over time. Figure 5 shows the average age of inventors rose by about 3 years between 2000 and 2016. A complementary measure, the share of inventors less than 36 years old, fell by almost 4 points over the same period. Among the entire civilian labor force, the share of workers less than 35 years old fell from 38.8 to 35.3 percent between 2000 and 2020. The share of young inventors after 2010 appears to be rising as the average age plateaus.

The aging inventor population might affect the nature and pace of innovation. Older inventors might bring experience and a nuanced understanding of problems. However, the risk-taking appetite and disruptive thinking associated with younger inventors might decline.

Fact 4: African Americans are significantly under represented among inventors, Asians inventors are increasingly common.
Our data provide a unique opportunity to summarize the self-reported race and ethnicity of inventors. Previous work aimed at describing the race of inventors relied on surname-based algorithms or matching curated lists of inventors (Cook and Kongcharoen, 2010; Cao et al., 2021; Breschi et al., 2017). We find substantial changes to the racial composition of inventors over this period. Figure 6 summarizes these changes by showing, for each race and ethnicity group, the ratio of the share of the inventor population in 2000 and 2016 to the group's share of the overall population. In both years the Black, American Indian or Alaska Native Alone (AIAN), Hispanics, and individuals with two or more race categories are under represented relative to their population share. The Black or African American Alone group, for instance, accounted for about 12.3 percent of the population in 2000 and only about 1.2 percent of inventors in 2000 (a ratio of 0.1 in the figure). Asians, on the other hand, had an inventor share in 2000 about 3.5 times

[16] See BLS Employment Projections, `https://www.bls.gov/emp/tables/civilian-labor-force-summary.htm` (accessed 09/2022).

Figure 5: Inventor Age



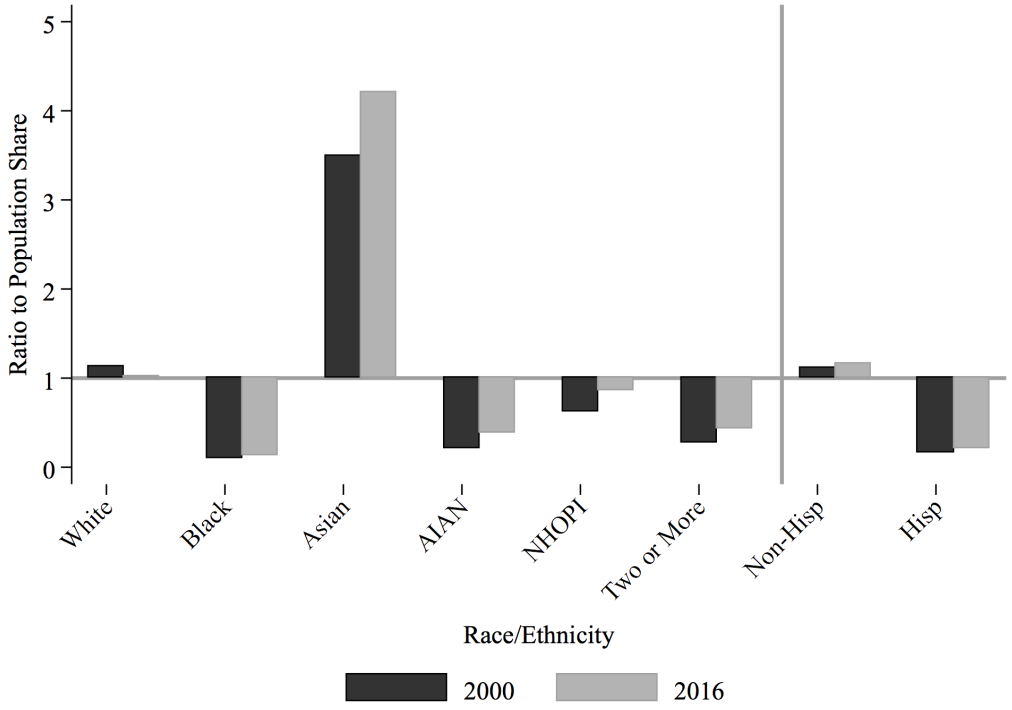
Source: Inventor Employment History
Notes: Inventor mean age is the mean age of all grant active inventors. Young inventor share is the share of grant active inventors ≤ 35.

their population share in 2000 (12.6% for inventors, 3.6% for 2000 Decennial). This rose to over 4.2 times by 2016 (21.9% for inventors, 5.2% for 2016 ACS).

In terms of a relative propensities, Figure 6 shows that the Native Hawaiian or Other Pacific Islander Alone (NHOPI) became more common among inventors from 2000 to 2016. To aid intuition about the racial and ethnic composition of inventors, we also consider the overall distribution of inventors across groups. Table 4 shows the inventor shares for each group in each period. We see that NHOPI accounted for about 0.06 percent of inventors in 2000 and rose to 0.17 percent by 2016. The increased representation of Asians is driven by a 9.3 percentage point increase (from 12.6 to 21.9). The share of the White Alone group fell by a similar amount, 10.7 percentage points (from 85.3 to 74.6). By ethnicity, the share of Hispanic inventors also rose from 2 to 3.7 percent, but remains well below the 2016 population share of 17.3 percent. The patterns we observe in the racial composition of inventors are consistent with previous studies showing the under representation of minority groups among inventors (Cook and Kongcharoen, 2010; Bell et al., 2019).[17]

[17]Bell et al. (2019) do not observe race and ethnicity in their tax data sample, but report inventor rates by race and ethnicity among the NYC school sample.

Figure 6: Inventor Relative Race and Ethnicity Composition



Source: Inventor Employment History, 2000 Decennial Census, 2016 American Community Survey
Notes: Ratio of inventor population share, among grant active inventors, in 2000 and 2016 to the 2000 Decennial Census and 2016 American Community Survey, respectively. AIAN stands for American Indian or Alaska Native and NHOPI stands for Native Hawaiian or Other Pacific Islander.

Underrepresentation can imply missed opportunities, both for the individuals and for society at large, in terms of economic contributions and social advancement. Our results may be driven by historical and systemic issues that might limit access for certain racial and ethnic groups. Educational disparities, network effects, and differential access to capital and resources play significant roles in an individual's career path. Such disparities can affect the scope and direction of innovation. Diverse teams are known to foster greater creativity and introduce a wider range of perspectives and ideas (Tsusaka et al., 2017).

Fact 5: Representation of different demographic groups varies significantly by sector. Having the employment histories of inventors allows us to simultaneously describe worker and employer characteristics. The demographic composition of the inventor population may be partly explained by variation in across sectors. Policy makers focused on patterns of representation would also need to account for the this type of heterogeneity.

Table 4: Change in Race/Ethnicity Composition

Race/Ethnicity	Population Share		Inventor Share	
	2000	2016	2000	2016
White Alone	75.1	73.3	85.26	74.63
Black or African American Alone	12.3	12.6	1.239	1.635
American Indian or Alaska Native Alone	.9	.8	.1882	.3074
Asian Alone	3.6	5.2	12.58	21.89
Native Hawaiian or Other Pacific Islander Alone	.1	.2	.06219	.1739
Two or More Races	2.4	3.1	.6678	1.363
Hispanic	12.5	17.3	2.002	3.692
Non-Hispanic	87.5	82.7	98	96.31

Source: Inventor Employment History, 2010 Decennial Census
Notes: Population shares are those from the 2000 Decennial Census and 2016 American Community Survey. Inventor shares for 2000 and 2016 are the share of grant active inventors in each year respectively.

Indeed, we find that the demographic composition of inventors varies significantly by sector. Table 5 shows the share of female, foreign born, young ($\leq$ 35), older ($\geq$ 56), and non-White inventors for a subset of sectors. The highest share of female inventors is in Health Care and Social Assistance (17.8%) and Educational Services (14.8%) and Finance, Insurance, Real Estate, Rental, and Leasing (14.3%). The inventor female share falls to 5 percent in Mining, Utilities, and Construction. The share of foreign born inventors is over 40 percent in Information and Educational Services. Information is also the sector with the highest share of young inventors (23%). Finally, non-White inventors are most common in Information (29.6%) and least likely to be in Mining, Utilities, and Construction (15.2%). These patterns demonstrate the importance of considering sectoral composition when describing inventor demographic characteristics.

The differences in demographic composition among inventors across various sectors highlight the importance of sector-specific dynamics in influencing innovation trends. Sectors like Health Care, Social Assistance, and Information Technology have higher proportions of female and foreign-born inventors, which may encourage a broader range of ideas and solutions. On the other hand, sectors like Mining, Utilities, and Construction, which show lower diversity in terms of demographics, might be missing out on these potential benefits. This could limit their capacity to innovate and adapt to new technologies or methodologies, which is crucial in today's rapidly changing technological landscape.

Table 5: Inventor Demographics by Sector

Sector	Female	Foreign Born	Age ≤ 35	Age ≥ 56	non-White
Mining, Util., Constr.	5.146	24.3	14.14	15.22	
Manufacturing	7.329	28.55	15.02	19.72	
Wholesale, Retail Trade	9.315	30.22	16.73	20.99	
Transp. & Warehousing	7.942	16.49	15.11	15.29	
Information	8.756	43.45	22.63	29.61	
FIRE	14.32	26.91	17.76	22.04	
Prof., Sci., & Tech. Services	9.489	38.54	15.64	26.06	
Educational Services	14.77	42.64	11.25	25.74	
Health Care & Social Ass.	17.8	31.31	10.96	20.15	
Other	12.39	24.69	14.97	16.97	
Overall	9.346	30.02	15.74	20.44	

Source: Inventor Employment History
Notes: Demographic shares of grant active inventors by sector. Female and Foreign Born are the female and foreign born share respectively. Age ≤ 35 and Age ≥ 56 are the share of grant active inventors aged less than or equal to 35 and greater than or equal to 56 respectively. non-White is the share of non-white inventors. FIRE denotes Finance, Insurance, Real Estate, and Rental and Leasing.

4.2 Employer Characteristics

Next, we turn to the characteristics of firms that employ inventors. Large firms engage in fundamentally different types of innovation than do smaller firms (Akcigit and Kerr, 2018). This has important implications for the allocation of R&D spending across firm types. This logic extends to inventors as well. The allocation of inventors across different types of firms can have important implications for aggregate innovative activity (Akcigit and Goldschlag, 2023). In this section, we describe the types of firms employing inventors. The data used for these analyses, in future work, could be used to decompose variation in inventor productivity by firm characteristics, explore the link inventor productivity and firm productivity, and the knowledge composition of industries based upon their inventor human capital.

Fact 6: Inventors work at older, larger firms.
We start with the distribution of inventors by firm age categories. Table 6 shows the distribution of inventor jobs by firm age and inventor age. Roughly 10 percent of inventor jobs are in young firms (≤ 5 years old). The majority of inventor jobs (68%) are in firms more than 20 years old. Consistent with the literature showing that young workers are over represented among young firms (Ouimet and Zarutskie, 2014), the share of inventor jobs at young firms is highest among young workers, nearly twice as large (15% and 8%)

among those 25 or younger compared to those 56 or older. These figures are similar to those for all workers found in the Quarterly Workforce Indicators, which shows that the percent of workers in young firms is roughly 15% for young workers (≤ 24) and 10% for older workers (≥ 55).

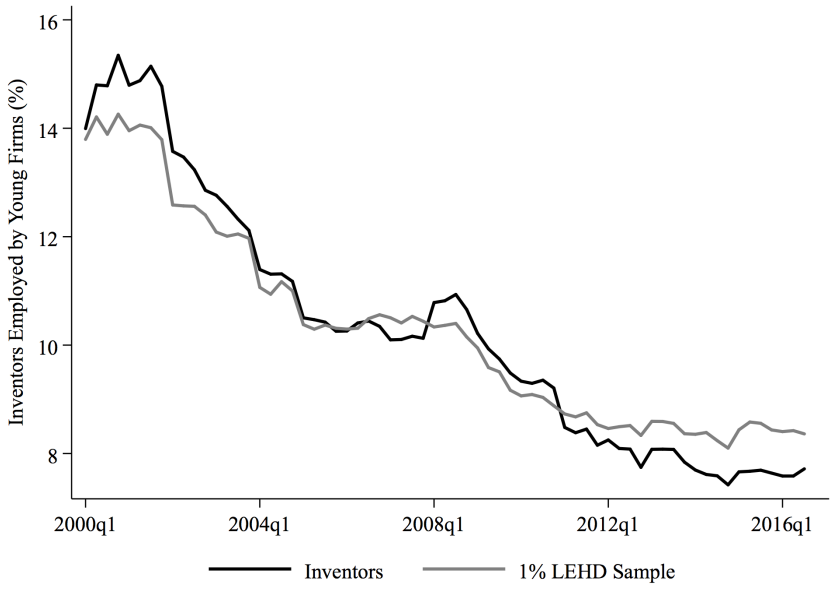
Table 6: Inventor Employer Firm Age Distribution by Inventor Age

Firm Age	All	≤ 25	$26-35$	$36-45$	$46-55$	$56+$
0 to 5	10.3	14.8	13	10.8	8.6	7.6
6 to 10	7.6	9	8.3	7.9	7	6.7
11 to 20	13.7	16.9	14.9	13.8	12.9	13.1
21+	68.3	59.3	63.7	67.5	71.5	72.7

Source: Inventor Employment History
Notes: Columns show the share of inventor-quarters for a given age group, or among all inventors, across firm age groups.

As shown in Figure 7, over our sample period the share of inventors at young firms declined from roughly 15 percent in the early 2000s to less than 8 percent by 2016.

Figure 7: Share of Inventors at Young Firms

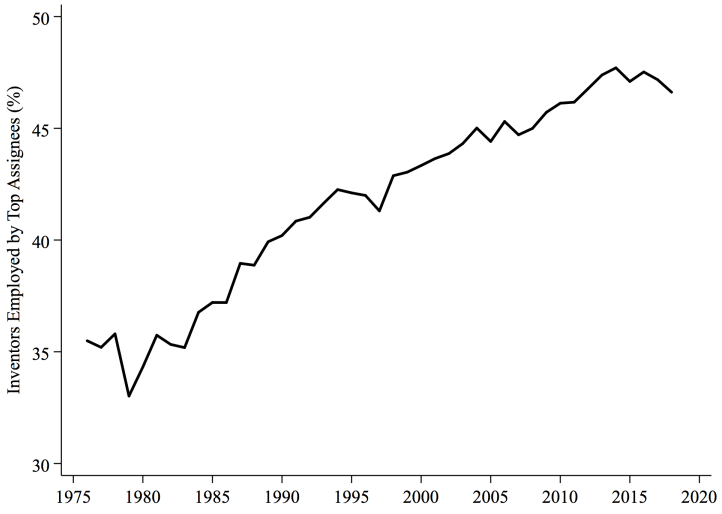


Source: Inventor Employment History
Notes: Figure shows the percent of inventors and workers in a 1% sample of the LEHD, capturing a representative sample of all workers, at young firms (≤ 5 years old). To make the 1% LEHD sample comparable to the inventor sample, it is re-weighted by the quarterly 3-digit NAICS composition of the inventor sample.

This pattern is not unique to inventors. Using a 1% sample of LEHD workers weighted by the industry composition of inventors, we see that the share of workers at young firms in industries similar to those of inventors fell from 14% to 8% between 2000 and 2016. Akcigit and Goldschlag (2023) show that the share of inventors in old, larger incumbent firms has risen substantially from 2000 to 2016.

This pattern of increasing concentration of inventors among incumbent firms appears to go back at least to the 1970s. Figure 8 shows the share of inventors in the PatentsView data on patent grants with an assignee that is among the top one percent of assignees by inventor count.[18] We find that the share of inventors on patent grants with the largest assignees between 1980 and 2018 rose by almost 13 percentage points from about 34 percent to 47 percent. Over that period, the count of unique inventors an assignee must be associated with to be classified as the top one percent rose from 81 to 154. These patterns suggest that the rising share of inventors employed by incumbents is a phenomenon that predates our sample period.

Figure 8: Share of Inventors on Grants Large Assignees



Source: PatentsView

Notes: Figure shows the percent of inventors (disambiguated using PatentsView inventor IDs) on patent grants each year that were linked, via a patent grant, to an assignee (disambiguated using PatentsView assignee IDs) that itself is among the top one percent of assignees by the number of associated unique inventors. The number of associated inventors necessary to be classified as among the top one percent is 81 in 1980 and rises to 154 in 2018.

[18] For this exercise we utilize PatentsView's disambiguated inventor and assignee identifiers. As shown in Table 3, the PatentsView inventor identifiers line up well with our PIK matches, an indication that the inventor disambiguation in PatentsView is high quality.

Compared to firm age, the distribution of inventors across firm size groups does not exhibit significant differences by inventor age. The gap between younger and older inventors in the propensity to work at small firms is much smaller than for young firms.[19] Table 7 shows the distribution of inventor jobs by firm size groups. About 10 percent of inventor jobs are at firms with 20 or fewer employees while almost 64 percent of inventor jobs are in firms with over 1000 employees. Inventors in the middle of the inventor age distribution, from 26 to 55 years old, are the most likely to work for the largest firms. This patterns are similar to what we see for the full workforce.[20]

Table 7: Inventor Employer Firm Size Distribution by Inventor Age

Firm Size	all	≤ 25	$26-35$	$36-45$	$46-55$	$56+$
1 to 20	10.1	13.7	9	9.3	10.3	12.9
21 to 250	16.7	20	16.7	16.3	16.2	17.9
251 to 1000	9.6	9.2	9.6	9.7	9.5	9.7
1000+	63.6	57.1	64.7	64.7	64	59.5

Source: Inventor Employment History
Notes: Columns show the share of inventor-quarters for a given age group, or among all inventors, across firm age groups.

In summary, the U.S. innovation ecosystem, and the utilization of inventive talent in particular, is dominated by large firms. However, smaller and younger firms play a crucial role in infusing the innovation ecosystem with essential vitality. These firms are often more agile and may outpace larger companies in innovation speed, despite having fewer inventors. A balanced ecosystem that includes both large and small firms can enhance competition, collaboration, and a diversity of innovative methods. Facilitating the growth of new startups and smaller firms through increased access to capital, mentorship programs, and university partnerships could lead to a rise in innovative job opportunities within these smaller firms.

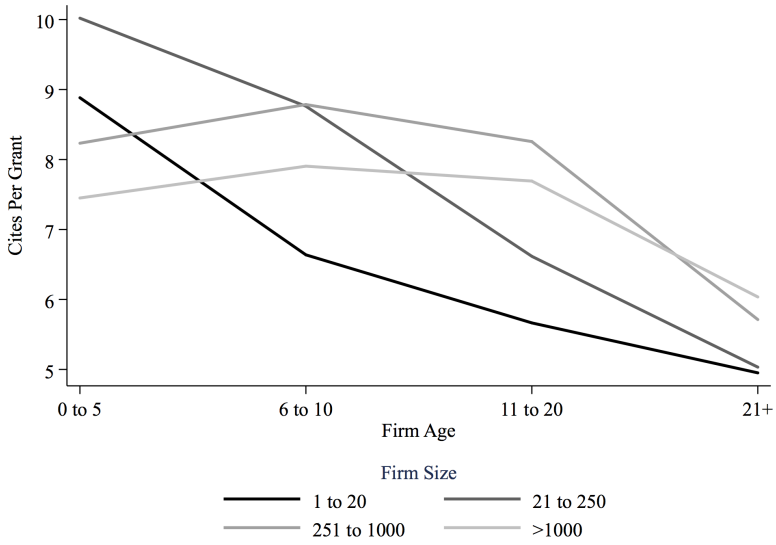
Fact 7: Inventors at young firms produce more impactful patents.
Inventor productivity varies significantly by firm age and firm size. Figure 9 shows inventor productivity, measured by average citations received per grant, by firm age

[19]The measure of firm employment is derived from the LBD, which captures full- and part-time employees, including salaried officers and executives of corporations, who are on the payroll in the pay period including March 12. Included are employees on paid sick leave, holidays, and vacations. We use LBD-LEHD linkages developed by Haltiwanger et al. (2014).

[20]In the QWI, we see that approximately 50% of individuals between 19 and 54 work for firms with 500+ employees.

and firm size groups of the inventor's employer. We see a monotonically negative relationship between firm age and inventor productivity among firms with 250 or fewer employees. Among larger firms, there is a gentle increase in inventor productivity to age 10, but falls thereafter. Inventors at old, small firms tend to have the lowest citations per patent grant. Inventors at very old, very large firms have citations per grant about 40 percent lower than those at mid-sized young firms (6.0 compared to 10.0). Except among the oldest group (21+), there is an inverse-U shape within firm age bins across firm size (vertical ordering of the lines). That is, within an age group, inventor productivity rises with firm size up to a point and then falls. The only exception to this is among the oldest group, where despite being lower in general, exhibits monotonically rising inventor productivity in size.

Figure 9: Firm Age, Firm Size, and Citations per Grant



Source: Inventor Employment History
Notes: Average citations per grant received within the first five years after grant by firm age and firm size groups averaged over time. Firm age groups: [0,5], [6,10], [11,20], and 20+. Firm size groups: [1, 20], (20, 250], (250,1000], 1000+.

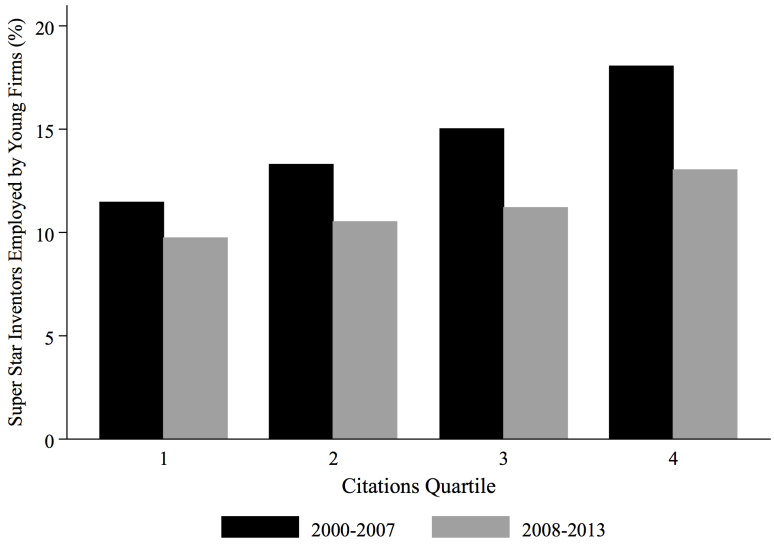
The observed inverse-U shape pattern within firm age groups across different firm sizes implies that as firms grow, their capacity for innovation increases up to a certain size but begins to decrease as they become very large. This might be due to increased complexity and rigidity in larger organizations. The exception among the oldest firms, where productivity monotonically rises with size, might reflect accumulated experience

and resources that help sustain innovation despite the firm's age.

Fact 8: Inventors, especially the most productive, are less likely to work at young firms.

Focusing on superstar inventors, those in the top 10 percent of the citations received distribution within an inventor cohort, we find that the share working for young firms declined more among inventors with the most impactful patents.[21] Figure 10 shows the change in the share of superstar inventors between the first and second half of our sample by the quality of their patent grants. Superstar inventors are placed into four equal size bins based upon the citations received of their patent grants. We see that the share working at young firms declined most for inventors with the most highly cited patents. There has been a marked flattening of the relationship between quality of inventive output and the likelihood of working at a young firm over this period.

FIGURE 10: SUPERSTAR INVENTORS AT YOUNG FIRMS, PRE & POST 2008



Source: Inventor Employment History

Notes: Figure shows the percent of superstar inventor jobs at young firms ($\leq$ 5 y.o.) in the first and second half of the sample by citation quartiles. Superstar inventors are those in the top 10 percent of the cumulative citations received distribution within cohort. Patent citations received quartiles split the sample of superstar inventors into four groups based upon the citations received of their granted patents.

[21] Our identification of superstar inventors is time varying. For each inventor, we compute the sum of cumulative 5-year windowed citations received up to and including the given quarter and divide by the mean of the same measure among all inventors of the same age in that quarter. Using a 5-year window of citations limits the impact of censoring near the end but also constrains the number of years. An inventor is identified as a superstar if they are in the top 10% of this quarter-age demeaned cumulative citation measure.

The decline in the share of superstar inventors at young firms might reflect challenges that these firms face in attracting and retaining top talent. This could be due to comparative disadvantages in compensation, resources, or market stability. It might also reflect a broader shift in the innovation ecosystem where venture capital and investments are becoming more risk-averse, favoring established players over newcomers.

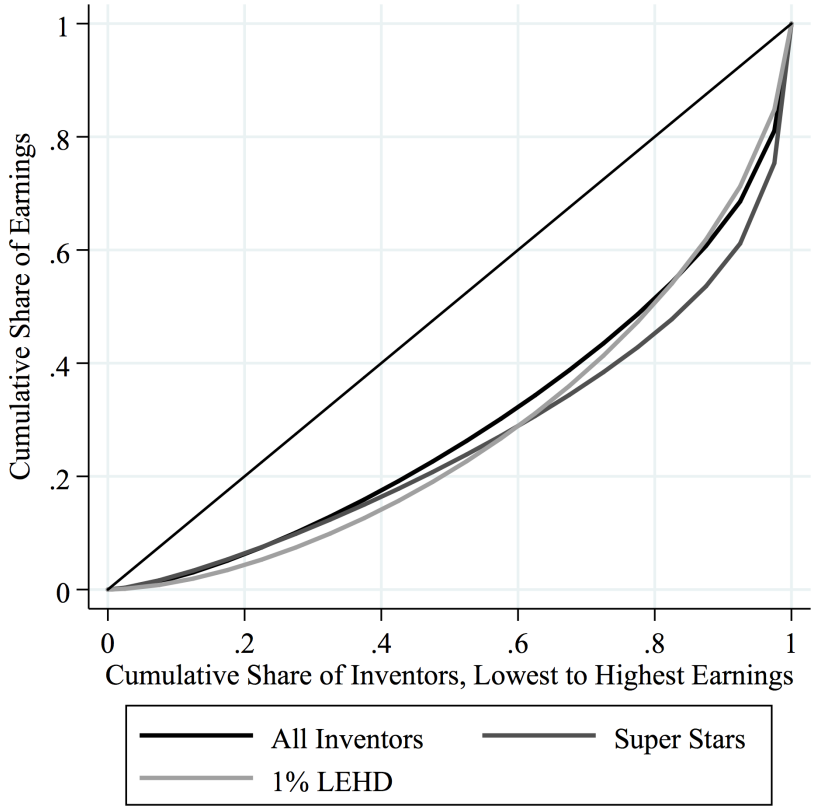
4.3 Earnings

Next we turn to the price mechanism that coordinates inventor and firm innovative activity—inventor earnings. There are a number of reasons inventor earnings might be related to aggregate innovative activity. An individual's innovative effort is responsive to incentives in both corporate and university settings (Lerner and Wulf, 2007; Lach and Schankerman, 2008). Moreover, evidence suggests that inventors employed by firms are able to capture rents associated with patent grants (Toivanen and Väänänen, 2012) and that taxation affects innovative effort (Akcigit et al., 2016, 2021). Firms may choose to share rents associated with a successful innovation in order to limit attrition among its high skill workers (Kline et al., 2019). Higher inventor compensation may also be a way for incumbent firms to attract talent away from entrants (Akcigit and Goldschlag, 2023). Moreover, the distribution of returns to high skilled labor, such as inventors, may also have implications for overall earnings inequality. In this section we describe the earnings of U.S. inventors, its skewness, and how earnings relates to inventive activity. Future research can extend these basic facts to investigate what firm and person characteristics explain variation in inventor earnings, the contribution of within and between firm variation, and the role of earnings in explaining the allocation of inventive talent between private and public institutions.

Fact 9: Inventor earnings is highly skewed, especially for superstar inventors.
We show the earnings distribution among inventors several ways. First, we examine the skewness of earnings among inventors relative to other workers. Figure 11 shows the Lorenz curve for inventors and a sample of all workers. The y-axis of the figure captures the cumulative share of earnings and the x-axis shows the cumulative share of individuals, ordered by earnings. Perfect earnings equality would place the curves on the 45 degree line. For the bottom half of the distribution, the curve for all workers is below the inventor curves, suggesting the bottom half of all workers accounts for less total earnings for all workers than the bottom half of inventors accounts for total earnings of inventors. The curve for inventors bows out at the top end of the plot, suggesting more high-end earnings inequality among inventors. This is especially true for super-

star inventors–the bottom 90 percent of superstar inventors account for 61 percent of superstar inventor earnings. In contrast, the bottom 90 percent of all inventors account for nearly 69 percent of inventor earnings and among all workers that figure is over 71 percent.

Figure 11: Inventor Earnings Lorenz



Source: Inventor Employment History

Notes: Figure shows the proportion of earnings assumed by the bottom x percent of the people. For the bottom x percent of inventors, what percentage (y percent) of the total earnings they have. The percentage of inventors is plotted on the x-axis, the percentage of earnings on the y-axis. The 1% LEHD sample is the employment history of a 1% sample of workers in the full LEHD data.

Next we provide a more direct comparison of inventor earnings to the earnings of all workers. Table 8 shows the percent of inventors and superstar inventors earning more than each percentile of earnings distribution of all workers. Inventors are over represented at the top of the earnings distribution. Over 92 percent of all inventors and almost 98 percent of superstar inventors earn more than the 60th percentile of the earnings distribution. The majority of inventors (63%) and the overwhelming majority

of superstar inventors (88%) are within the top decile of the earnings distribution. At the upper tail of the distribution, the top 1 percent, almost 8 percent of inventors and 19 percent of superstar inventors earn more than 99 percent of the population.

Table 8: Inventor Relative Wage Distribution

Population Earnings Percentile	Percent of Inventors Above	Percent of Superstar Inventors Above
99	7.71	18.77
98	15.69	37.27
96	32.03	64.5
94	45.3	77.87
92	55.26	84.6
90	62.69	88.35
80	81.47	95.225
60	92.347	97.898

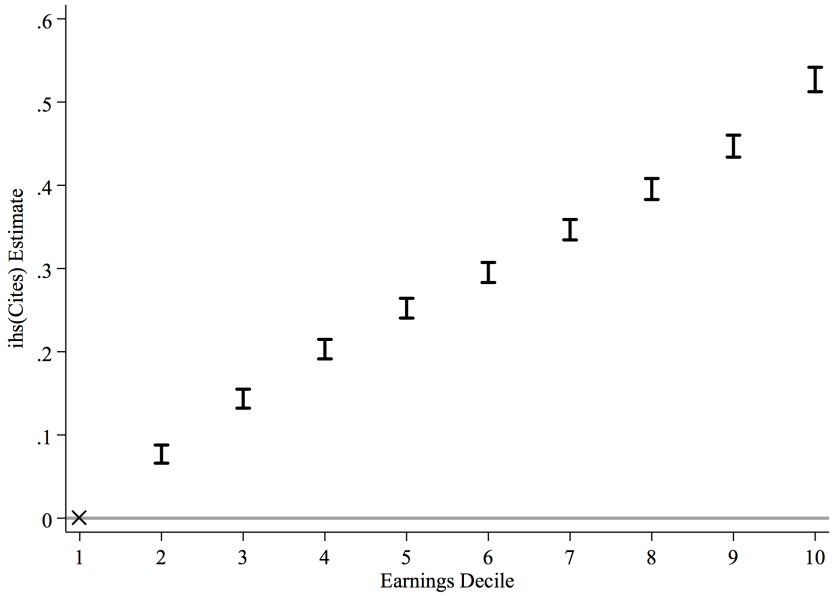
Source: Inventor Employment History
Notes: Table shows the percent of inventors with earnings less than each percentile of earnings distribution for the population. The population is approximated using a 1 percent sample of the LEHD data with a constant state composition. Earnings percentiles and the share of inventors with earnings less than each percentile are calculated within each year-quarter from 2000q1 to 2016q3. This table shows the mean shares across quarters.

Overall, these findings highlight the significant financial advantages that successful inventors can gain, reflecting both the high premium placed on innovation and the potential for considerable income disparities among inventors. This demonstrates that while innovation can be financially lucrative, the rewards tend to be unevenly distributed, with the majority of benefits accruing to a select group of highly successful individuals. Such substantial financial incentives can attract highly skilled and talented individuals to fields driven by innovation, potentially enhancing U.S. innovation capacity by drawing in bright minds from less innovative sectors. However, the pronounced skew towards high earnings for top inventors might also deter skilled individuals who do not see a clear path to achieving top-tier status, potentially restricting the diversity of talent engaged in innovative activities.

Fact 10: Inventor earnings is closely tied to inventive productivity.
Consistent with the literature showing that inventor effort is responsive to earnings, we find a strong positive relationship between earnings and the quality of inventive output.

To examine this, we estimate a simple regression of earnings on citations received. Earnings is measured as the average within a window around the current quarter to limit the impact of transient earnings shocks around the time of a patent grant. We segment inventor jobs into deciles of the windowed earnings distribution and control for age, along with year, industry, firm age, and firm size fixed effects. Estimates are shown in Figure 12. We find a positive and monotonic relationship between earnings and citation weighted patent counts. Relative to inventors in the bottom earnings decile, inventors in the top decile produce patents with over 50 percent more citations. Though causality likely runs in both directions, these partial correlations are consistent with the literature showing the financial returns to inventors of their innovative activity (Toivanen and Väänänen, 2012; Kline et al., 2019).

Figure 12: Earnings and Citations



Source: Inventor Employment History
Notes: This figure shows estimates of λ_j, with $j = 1$ being the excluded group, from the following equation:

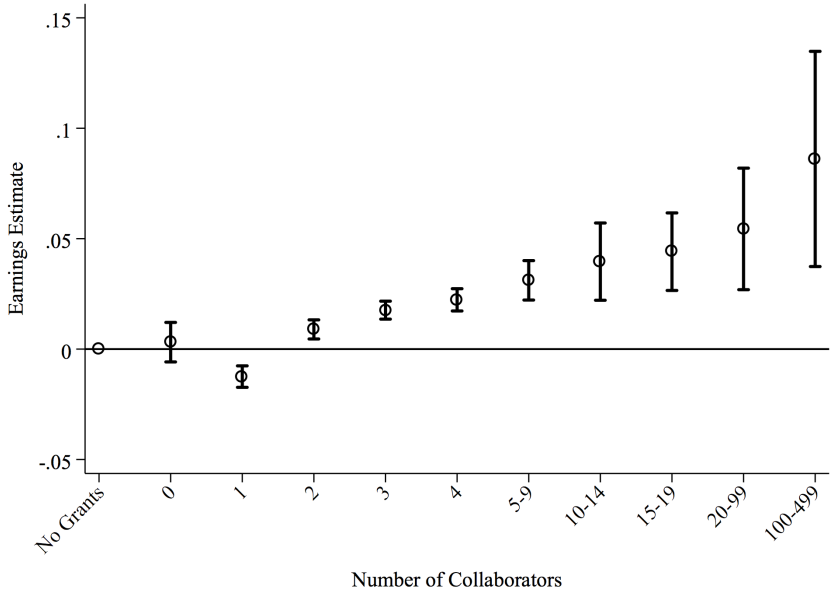
$$ihs(Cites_{i,t}) = \alpha + \sum_{j=1}^{10} \lambda_j EarnL2F2[j]_{i,t} + \boldsymbol{X} + \epsilon_{i,t} \quad (1)$$

ihsCites is the ihs() transformed 5-year window cites of patents granted in t. $EarnL2F2$ is the decile bin of average earnings in $t-2$ to $t+2$. $\boldsymbol{X}$ contains age and age squared, year FE, 4-digit NAICS FE, firm age group FE, and firm size group FE. Clustered at the individual level.

Finally, invention has increasingly become a team exercise. The average number of

inventors on a patent document, or indeed scientists on published papers, has risen over time (Wuchty et al., 2007). Economic theories emphasize the crucial role of knowledge spillovers in fostering growth, suggesting that interactions and partnerships among inventors–through co-invention networks–can enhance the flow of knowledge and boost productivity (Romer, 1990; Aghion and Howitt, 1992; Azoulay et al., 2010; Akcigit et al., 2018). These collaborative networks may also lead to higher compensation for inventors who engage with more collaborators (Lucas Jr and Moll, 2014; Perla and Tonetti, 2014). Figure 13 shows the relationship between an inventor's earnings and their number of unique co-inventors.

Figure 13: Earnings and Collaborators



Source: Inventor Employment History
Notes: This figure shows estimates of λ_j the following equation:

$$ln(earn_{i,t}) = \alpha + \sum_{j=1}^{12} \lambda_j collabBin[j]_{i,t} + \boldsymbol{X} + \epsilon_{i,t} \quad (2)$$

$ln(earn_{i,t})$ is full quarter log real earnings. *collabBin* are bins of the number of unique, cumulative co-inventors of inventor i at time t. The base bin of *collabBin*, excluded from the regression, captures inventors that do not yet have a grant. X contains controls including year, 4-digit NAICS, firm size bin, firm age bin, and person fixed effects.

Lone (0 collaborators) inventors appear to have similar earnings to inventors that have not yet had their first patent grant (the "No Grant" category). Inventors with a

single collaborator earn slightly less, but at two or more collaborators the relationship between collaborators and earnings rises monotonically. The patterns in Figure 13 are consistent with the view that collaboration is a driver of inventor productivity.

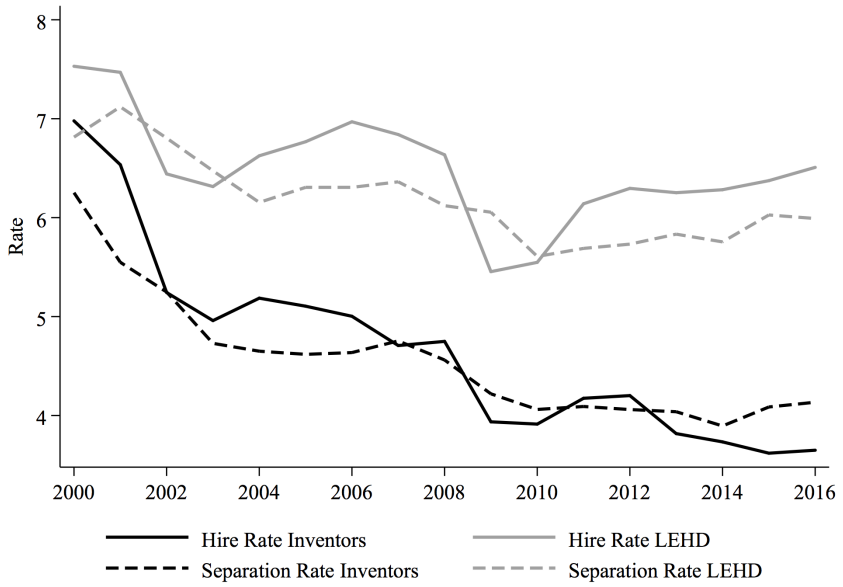
The growing trend toward collaboration, illustrated by the rise in the average number of inventors per patent, aligns with economic models that emphasize the benefits of knowledge spillovers for enhancing productivity and fostering innovation. Collaborative efforts allow inventors to merge diverse skills, exchange knowledge more efficiently, and create novel ideas that are unlikely to emerge in solitary environments, leading to richer and more inventive outcomes. Additionally, the observation that inventors with more collaborators often receive higher compensation suggests that such cooperative efforts are not only effective but also highly valued.

4.4 Employment Dynamics

Finally, we turn to the employment dynamism of inventors. Numerous measures of employment dynamics exhibit trend declines in recent decades. Declines appear in measures of job creation and destruction from firm and establishment data, job flows from survey data, and hires, separations, and job-to-job flows from administrative employee employer data (Davis et al., 2006, 2012; Lazear and Spletzer, 2012; Hyatt and Spletzer, 2013). The reallocation of inventors from unproductive firms to productive ones may be especially important for aggregate economic performance and productivity growth. In this section we measure the rate at which inventors change jobs, start new firms, and relocate across geographies. Our inventor jobs data offer especially rich potential in understanding the employment dynamics of inventors. Future work, using these data, can be used to explore spin-offs, how the movement of inventors affects the flow of knowledge across firms and industries, and the creation and destruction of inventor teams within and across firms.

Fact 11: Inventors are less likely to switch jobs over time.
Inventors saw a significant decline in hire and separation rates between 2000 and 2016. Figure 14 shows the hire and separation rates for inventors and a sample of all workers weighted by the industry composition of the inventor sample. Hire and separation rates for inventors remained below that of the overall workforce over the entire period. The overall work force also saw a decline in both the rate of hires and separations but both rates stabilize after 2010. Inventors, on the other hand, exhibit a decline in hire rates through the end of the sample period and a separation rate that stabilizes after 2010 but at a much lower level than in the early 2000s.

Figure 14: Inventor Hire and Separation Rates



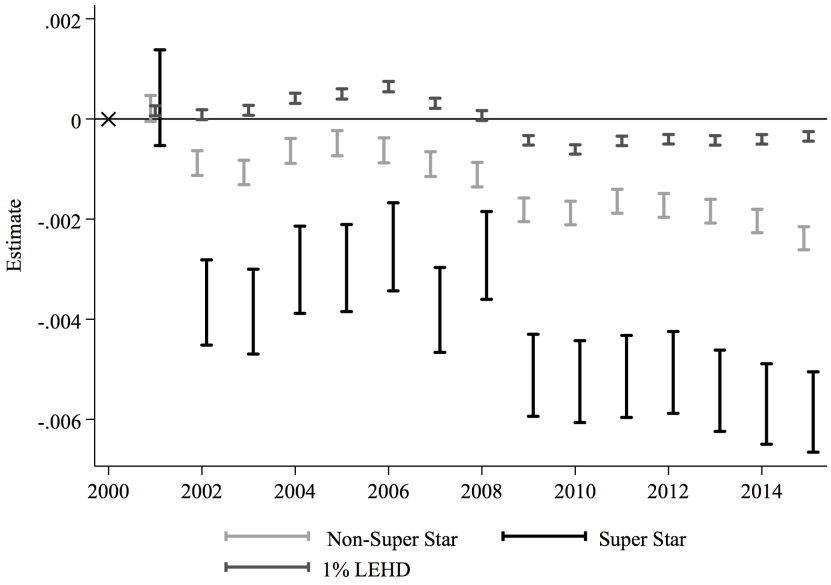
Source: Inventor Employment History
Notes: Figure shows Hire and Separation (QWI definition of stable hires and separations) for inventors and a 1% sample of the LEHD for comparison. The 1% LEHD sample is weighted to reflect the time varying industry composition of the inventor sample and thus captures the hire and separation of workers in industries in which inventors are most frequent.

Lower turnover among inventors might suggest a more stagnant job market, which could lead to reduced dynamism in the innovation sector. Frequent movement of talent between firms is often associated with vibrant knowledge exchange and innovation. If inventors stay within the same company or environment for longer periods without new challenges or exposures, it might limit the cross-pollination of ideas that can spur new inventions.

Fact 12: Inventors, especially superstar inventors, are less likely to start a firm.
Another measure of economic dynamism among inventors is the rate at which they start new firms. The rate at which new businesses have started in the U.S. has declined for decades (Decker et al., 2014). One might hope that despite this more persistent, general decline in entrepreneurship the number of high quality entrants remains robust (Guzman and Stern, 2020). Indeed, Miranda and Zolas (2017) find that among non-employers, inventors are more than twice as likely than non-inventors to transition to employer businesses. We find, however, that inventors are less likely to become entrepreneurs over time. Figure 15 shows estimates of the change in the likelihood of an

inventor becoming an entrepreneur.

Figure 15: Change in Inventor Entrepreneurship Rate



Source: Inventor Employment History
Notes: This figure shows estimates from the following:

$$Entrep_{i,t} = \alpha + \sum_{k=2001}^{2015} \beta_k D^k + \psi_i + \epsilon_{i,t} \tag{3}$$

Where $Entrep_{i,t}$ is an indicator of whether an inventor, or an individual in the 1% LEHD sample, i is a startup founder in year t. D^k are year dummies, and ψ_i are inventor fixed effects. For corporations, an inventor will be flagged as a founder of the startup if they have positive earnings in the first quarter of a firm's operations and are among the top three earning workers in the firm's first year. For sole proprietorships, an inventor will be flagged as an entrepreneur if they appear as the business owner in tax filings or are among the top two earners in the firm's first year. Partnerships are excluded from the firm founder database because business owners are prohibited from paying themselves wages that would appear in the LEHD data. See Choi et al. (2021) for additional measurement details.

Specifically, the figure shows differences in the rate of entrepreneurship, relative to 2000, for superstar inventors, non-superstar inventors, and the 1% sample of the LEHD, controlling for individual fixed effects. The share of inventors being a firm founder each year is quite low. In 2000, about 0.64 percent of inventors started a business, including about 1 percent of superstar inventors and 0.59 percent of non-superstar inventors. Among the 1% LEHD sample only about 0.29% started a business in 2000. The estimates

suggest that the already low rate of entrepreneurship among inventors fell significantly. Our estimates imply that by 2015, the rate at which non-superstar inventors started businesses fell by 41 percent and for superstar inventors the decline was even steeper, falling by 57 percent. These declines are much larger than we see for the LEHD sample. Indeed, in the early 2000s the entrepreneurship rate rises for the 1% LEHD sample and then falls after 2006. This mirrors the overall trend in startup rates, which rose from 9.1% in 2000 to 10.3% in 2006 then falls to 7.6% in 2010.[22]

Not only does the rate at which inventors start businesses matter for innovation and growth, but also the characteristics of those businesses. Indeed, as shown in Akcigit and Goldschlag (2023), inventor founded firms have more employees at startup and have as steeper life cycle profile, growing faster than non-inventor founded firms. Even so, it could be that the inventor founded firms that enter later in the sample are of significantly higher quality, making up for the fact that there are fewer of them. To assess this, we compute the average employment growth, conditional on survival, for each startup in its first five years after entry. We then compute several moments of the average post-entry growth rate distribution for inventor founded firms and non-inventor founded firms. Figure 16 shows the 90^{th}, mean, and 10^{th} percentile of post-entry growth rates by cohort and whether the startup has an inventor founder. The figure suggests that the quality of inventor founded firms, as measured by post-entry growth, has been fairly stable over time. Consistent with the steeper size-age profiles shown in Akcigit and Goldschlag (2023), inventor founded firms average and 90^{th} percentile growth rates are higher than non-inventor founded firms.

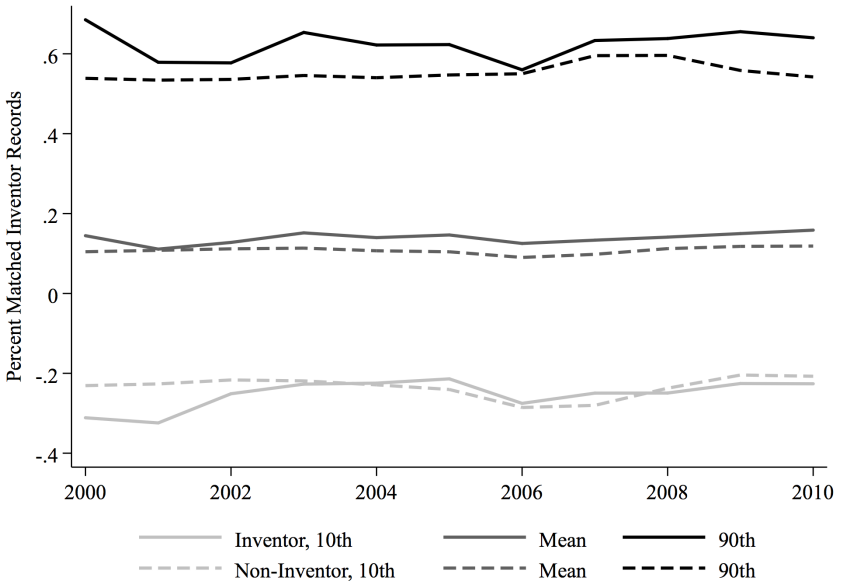
The rapid growth and larger initial size of firms founded by inventors suggest that these businesses may exert a greater economic influence compared to those started by non-inventors. This advantage could lead to increased job creation and significant contributions to GDP, especially within technologically advanced and innovation-heavy sectors. Furthermore, the rapid expansion and scalability of inventor-founded firms indicate a potentially enhanced pace of innovation. These companies are likely to introduce novel products and technologies or refine existing ones more swiftly, thus securing a competitive edge and becoming market leaders.

Fact 13: Inventors are increasingly geographically concentrated and less likely to change employment across state lines.

Finally, we explore the geographic distribution of inventors. The geographic distribution of inventive activity may have significant economic consequences if local exposure

[22]See the Census Bureau's Business Dynamics Statistics for details.

Figure 16: Startup Growth Rates

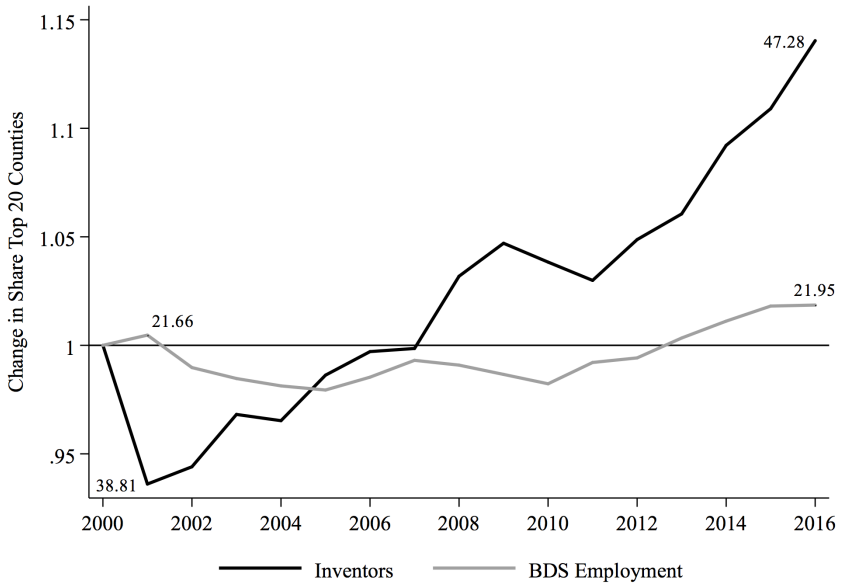


Source: Inventor Employment History
Notes: Figure shows the 90^{th}, 10^{th} percentiles, and average post-entry growth rates, conditional on survival, averaged over the first five years after entry. Statistics are computed for inventor-founded firms and non-inventor founded firms. Inventor-founded firms are identified as in Choi et al. (2021).

to inventors affects an individual's propensity to become an inventor (Bell et al., 2019). Consistent with the literature on agglomeration and the geographic clustering if innovative activity, we find significant geographic concentration of U.S. inventors. Indeed, the top 20 counties by number of inventors accounted for about 39 percent of inventors in 2001 but over 47 percent of inventors in 2016. Figure 17 shows the change in the share of inventors in the 20 largest counties by inventor count over time. The figure shows a similar measure for all workers—the share of all workers accounted for by the 20 largest counties by employment. For inventors, by 2016 the share of all inventors in those top 20 counties rose by nearly 14 percent (6 percentage points). We do not see a similar rise in concentration among all workers, which rose modestly from 21.6 percent in 2000 to 22 percent in 2016. These patterns are consistent with papers documenting the rising concentration of patenting activity in large urban and tech centers (Bettencourt et al., 2007; Chattergoon and Kerr, 2022).

Relatedly, we find that inventors are also less likely to change employment across-state boundaries. This might be expected given the decades long decline in mobility rates, a pattern that has been related to declining labor market fluidity (Molloy et al.,

Figure 17: Change in Top 20 County Concentration

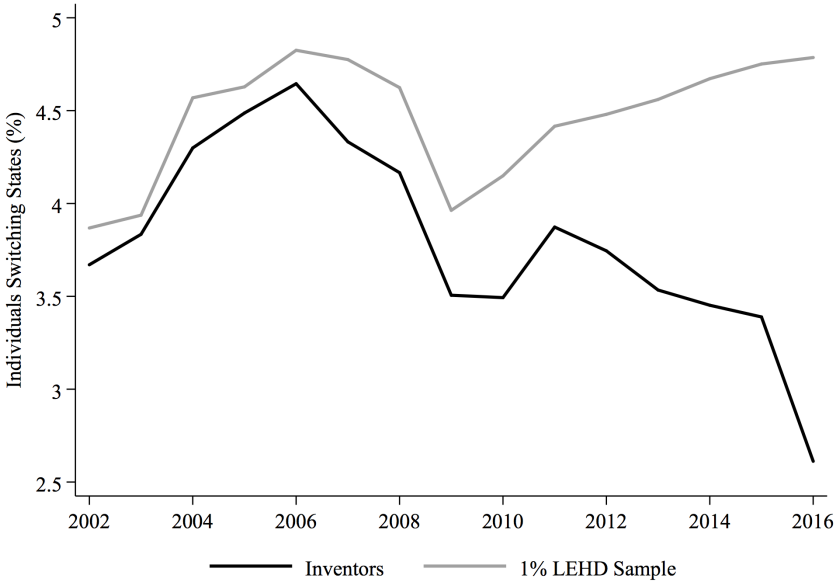


Source: Inventor Employment History, Business Dynamics Statistics (2018)
Notes: Figure shows change in share of grant active inventors by county accounted for by the top 20 counties by the count of grant active inventors. Change is measured as the ratio shr_t / shr_{2000}. Note that inventors are allowed to contribute to multiple counties in a year. The BDS employment change is measured analogously as the share of employment accounted for by the 20 largest counties by employment, which are not necessarily the same counties as the top counties for inventors. The set of counties is time varying, though there is relatively little volatility in the set of counties over time. The labeled points show the ending share in 2016.

2014). Figure 18 shows the share of inventors and the share of LEHD workers that change employment across state boundaries each year. Among all workers, this share rose in the early and mid 2000s, declined during the Great Recession, then rose again from 2009 to 2016. Inventors largely follow a similar pattern through the Great Recession, but did not experience a similar rise in the mid 2010s. Inventors are increasingly less likely to change employment across state lines when compared to the typical worker.

Taken together, these results suggest that the labor market outcomes of inventors are becoming less dynamic over time. Inventors are less likely to change jobs, less likely to start a new firm, increasingly concentrated in relatively few locations, and less likely to change employment across state boundaries. This shift could impact the spread of innovation across the U.S. and might necessitate a reevaluation of strategies to encourage more widespread and diverse innovative activities.

FIGURE 18: STATE MOVERS



Source: Inventor Employment History
Notes: Share of all (not just grant active) inventors and 1% LEHD sample workers that changed states at least once for their dominant BoQ job in a year.

5 Conclusion

There has long been interest in the characteristics of inventors and their role in the innovation process. Only recently have researchers begun to link inventors in patent documents to large-scale survey, census, and administrative data sources. Those linkages dramatically expand the types of analyses that can be done to open the "black box" of innovative activity. In this paper, we describe a novel effort to link patent inventors to Census microdata on individuals and their employment outcomes. We describe a number of patterns in the data that are relevant to both policy makers and researchers interested in the role of inventors in the economy.

We organize our analyses around two critical components of the U.S. innovation ecosystem: inventors, their employers, and their interactions in the labor market. In our first set of empirical results, focused on who becomes an inventor, we find that over 30% of inventors were born outside the U.S., the population of inventors as a whole is aging, African Americans are significantly under represented among inventors and the share of Asians among inventors nearly doubled from 2000 to 2016. We also find, consistent with prior work, that females are underrepresented among inventors, accounting for

less than 12% of inventors. Understanding the characteristics of inventors is crucial for identifying patterns of representation in the inventor talent pipeline, the allocation of talent, and ultimately the innovative capacity of the economy.

Our second set of basic facts describe innovative firms, which have an outsized impact on innovation and growth. We find that inventors are concentrated in older, larger firms. Almost 63% of inventors work in firms with more than 1,000 employees that are over 20 years old. Inventors at young firms tend to produce more impactful patents, but highly productive inventors are increasingly less likely to work in young firms over time. These changing compositional patterns—inventors working for older, larger firms—combined with cross sectional patterns of inventor productivity—inventors at young firms are more productive—raise important questions about the allocation of inventive talent.

In our third set of empirical results, we focus on inventor earnings. We find that inventor earnings is highly skewed, with 63% of inventors falling in the top 10% of the earnings distribution and 19% of superstar inventors falling in the top 1 percent of the earnings distribution. Earnings is highly correlated with inventor output and inventors with more collaborators have higher earnings. Characterizing inventor earnings illuminates the incentives inventors face and is therefore highly relevant for innovation and tax policy.

In our final section of empirical results, we focus on the employment dynamics of inventors. We find declining employment dynamism among inventors. The hire and separation rate among inventors fell over time and inventors are less likely to become entrepreneurs over time. Inventors are also becoming increasingly geographically concentrated, and are less likely to move across state lines. Employment dynamism plays an important role in reallocating resources between firms and geographic regions. Declining trends in dynamism among inventors could signal a weakening of reallocation pressures, which in turn could affect the innovative capacity of the economy.

The analyses and insights in this paper are relevant for several types of policies.

- *Encouraging Diversity in Innovation:* Educational programs and professional development initiatives such as targeted scholarships, mentorship programs, and community outreach efforts might lower barriers to entry in STEM fields, particularly for underrepresented groups.

- *Support Across Generations:* Education reforms, continuous learning opportunities, and policies that focus on incentives facing young entrepreneurs and integrating expertise of older workers with creative young talent may simultaneously address declining inventor entrepreneurship and the aging inventor workforce.

- *Support for SMEs:* Policies designed to aid small and medium-sized enterprises in overcoming innovation barriers such as tax incentives, subsidies for research and development and streamlined regulatory processes could spur the creation of new companies and support young firms creating high-impact innovations.

- *Fostering Mobility and Collaboration:* Enhancing the mobility of inventors and facilitating their collaboration through networks and shared spaces can help maintain an ecosystem that is conducive to high-impact innovation.

Observing these patterns is only possible because of the novel data linkages we introduce in this paper. Linking patent inventors to the survey, census, and administrative data housed at the Census Bureau opens a wealth of research opportunities. These linkages are available to researchers with approved access via the Federal Statistical Research Data Center (FSRDC) system.[23] We hope that by making these links available to qualified researchers on approved projects, the scope of questions that can be asked and answered about inventors and innovative activity will increase considerably.

[23]See `https://www.census.gov/about/adrm/fsrdc.html` for details about the FSRDC system and how to apply for access to confidential data (accessed 09/2022).

References

Abowd, John M, Kevin L McKinney, and Nellie L Zhao (2018) "Earnings inequality and mobility trends in the United States: Nationally representative estimates from longitudinally linked employer-employee data," *Journal of Labor Economics*, 36 (S1), S183–S300.

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen (2017) "The Social Origins of Inventors," National Bureau of Economic Research Working Paper 24110.

——— (2018) "On the Returns to Invention within Firms: Evidence from Finland," in *AEA Papers and Proceedings*, 108, 208–12.

——— (2023) "Parental Education and Invention: The Finnish Enigma," *International Economic Review*, 64 (2), 453–490.

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

Akcigit, Ufuk and Sina T Ates (2021) "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory," *American Economic Journal: Macroeconomics*, 13 (1), 257–298.

——— (2023) "What Happened to US Business Dynamism?" *Journal of Political Economy*, 131 (8), 2059–2124.

Akcigit, Ufuk, Salomé Baslandze, and Stefanie Stantcheva (2016) "Taxation and the International Mobility of Inventors," *American Economic Review*, 106 (10), 2930–81.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi (2018) "Dancing with the Stars: Innovation through Interactions," National Bureau of Economic Research Working Paper Nr 24466.

Akcigit, Ufuk and Nathan Goldschlag (2023) "Where Have All the "Creative Talents" Gone? Employment Dynamics of US Inventors," National Bureau of Economic Research Working Paper Nr. 31085.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas (2017) "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age," National Bureau of Economic Research Working Paper 23137.

Akcigit, Ufuk, John Grigsby, Tom Nicholas, and Stefanie Stantcheva (2021) "Taxation and Innovation in the Twentieth Century," *Quarterly Journal of Economics*, 137 (1), 329–385.

Akcigit, Ufuk and R. William Kerr (2018) "Growth through Heterogeneous Innovations," *Journal of Political Economy*, 126 (4), 1374–1443.

Akcigit, Ufuk, Jeremy Pearce, and Marta Prato (2024) "Tapping into Talent: Coupling Education and Innovation Policies for Economic Growth," *Review of Economic Studies*, forthcoming.

Andrews, Dan, Chiara Criscuolo, and Peter N. Gal (2016) "The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy," OECD Productivity Working Paper 5/2016.

Anonymous (1948) "The Eternal Apprentice," *Time Magazine*.

Argente, David, Salomé Baslandze, Douglas Hanley, and Sara Moreira (2020) "Patents to Products: Product Innovation and Firm Dynamics," CEPR Discussion Paper No. DP14692.

Azoulay, Pierre, Joshua S Graff Zivin, and Jialan Wang (2010) "Superstar Extinction," *Quarterly Journal of Economics*, 125 (2), 549–589.

Azoulay, Pierre, Benjamin F Jones, J Daniel Kim, and Javier Miranda (2020) "Age and High-growth Entrepreneurship," *American Economic Review: Insights*, 2 (1), 65–82.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen (2019) "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *Quarterly Journal of Economics*, 134 (2), 647–713.

Bessen, James (2008) "The Value of US Patents by Owner and Patent Characteristics," *Research Policy*, 37 (5), 932–945.

Bettencourt, Luis MA, Jose Lobo, and Deborah Strumsky (2007) "Invention in the City: Increasing Returns to Patenting as a Scaling Function of Metropolitan Size," *Research Policy*, 36 (1), 107–120.

Breschi, Stefano, Francesco Lissoni, and Gianluca Tarasconi (2017) "Inventor data for research on migration and innovation: The ethnic-inv pilot database," in *The International Mobility of Talent and Innovation: New Evidence and Policy Implications*: Cambridge University Press.

Cao, Chunfang, Xiaohui Li, Xiaoyang Li, Cheng Zeng, and Xuan Zhou (2021) "Diversity and Inclusion: Evidence from Corporate Inventors," *Journal of Empirical Finance*, 64, 295–316.

Celik, Murat Alp (2023) "Does the Cream Always Rise to the Top? The Misallocation of Talent and Innovation," *Journal of Monetary Economics*, 133, 105–128.

Chattergoon, Brad and William R Kerr (2022) "Winner Takes All? Tech Clusters, Population Centers, and the Spatial Transformation of US Invention," *Research Policy*, 51 (2), 104418.

Choi, Joonkyu, Nathan Goldschlag, John C Haltiwanger, and J Daniel Kim (2021) "Founding Teams and Startup Performance," National Bureau of Economic Research Working Paper 28417.

Chow, Melissa C, Teresa C Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Ruth Perlman, Martha Stinson, and T Kirk White (2021) "Redesigning the Longitudinal Business Database," National Bureau of Economic Research Working Paper Nr 28839.

Cohen, Wesley M, Richard Nelson, and John P Walsh (2000) "Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (Or Not)," National Bureau of Economic Research Cambridge Working Paper Nr 7552.

Cook, Lisa D and Chaleampong Kongcharoen (2010) "The Idea Gap in Pink and Black," National Bureau of Economic Research Working Paper Nr. 16331.

Davis, Steven J, R Jason Faberman, and John Haltiwanger (2012) "Labor market flows in the cross section and over time," *Journal of Monetary Economics*, 59 (1), 1–18.

Davis, Steven J, John Haltiwanger, Ron Jarmin, and Javier Miranda (2006) "Volatility and dispersion in business growth rates: Publicly traded versus privately held firms," in Acemoglu, Daron, Kenneth Rogoff, and Michael Woodford eds. *NBER Macroeconomics Annual*, 107–179: MIT Press.

Decker, Ryan A, John Haltiwanger, Ron S Jarmin, and Javier Miranda (2020) "Changing Business Dynamism and Productivity: Shocks versus Responsiveness," *American Economic Review*, 110 (12), 3952–3990.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda (2014) "The Role of Entrepreneurship in U.S. Job Creation and Economic Dynamism," *Journal of Economic Perspectives*, 28 (3), 3–24.

Dreisigmeyer, David, Nathan Goldschlag, Marina Krylova, Wei Ouyang, and Elisibeth Perlman (2018) "Building a Better Bridge: Improving Patent Assignee-Firm Links,"Technical report, Center for Economic Studies CES-TN-2018-01.

Graham, Stuart JH, Cheryl Grim, Tariqul Islam, Alan C Marco, and Javier Miranda (2018) "Business Dynamics of Innovating Firms: Linking US Patents with Administrative Data on Workers and Firms," *Journal of Economics & Management Strategy*, 27 (3), 372–402.

Griliches, Zvi (1992) "The search for R&D spillovers," *Scandinavian Journal of Economics*, 94, 29–47.

——— (1998) "Patent statistics as economic indicators: a survey," in *R&D and productivity: the econometric evidence*, 287–343: University of Chicago Press.

Guzman, Jorge and Scott Stern (2020) "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014," *American Economic Journal: Economic Policy*, 12 (4), 212–43.

Hall, Bronwyn (1996) "The private and social returns to research and development," in *Technology, R&D, and the Economy*, 140–183: Brookings Institution and the American Enterprise Institute.

Haltiwanger, John, Henry R Hyatt, Erika McEntarfer, Liliana Sousa, and Stephen Tibbets (2014) "Firm Age and Size in the Longitudinal Employer-household Dynamics Data,"Technical report, Center for Economic Studies CES-WP-14-16.

Haltiwanger, John, Ron S Jarmin, Robert Kulick, and Javier Miranda (2016) "High growth young firms: contribution to job, output, and productivity growth," in *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*: University of Chicago Press.

Henderson, Cody and Henry Hyatt (2012) "Estimation of Job-to-Job Flow Rates Under Partially Missing Geography,"Technical report, Center for Economic Studies CES-12-29.

Hopenhayn, Hugo and Richard Rogerson (1993) "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, 101 (5), 915–938.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow (2019) "The Allocation of Talent and US Economic Growth," *Econometrica*, 87 (5), 1439–1474.

Hsieh, Chang-Tai and Peter J Klenow (2009) "Misallocation and manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124 (4), 1403–1448.

Hyatt, Henry R and James R Spletzer (2013) "The recent decline in employment dynamics," *IZA Journal of Labor Economics*, 2 (1), 5.

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

Jarmin, Ron S and Javier Miranda (2002) "The longitudinal business database."

Jones, Benjamin F (2010) "Age and Great Invention," *Review of Economics and Statistics*, 92 (1), 1–14.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar (2019) "Who profits from patents? rent-sharing at innovative firms," *Quarterly Journal of Economics*, 134 (3), 1343–1404.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017) "Technological Innovation, Resource Allocation, and Growth*," *Quarterly Journal of Economics*, 132 (2), 665–712, 10.1093/qje/qjw040.

Lach, Saul and Mark Schankerman (2008) "Incentives and invention in universities," *RAND Journal of Economics*, 39 (2), 403–433.

Lazear, Edward P and James R Spletzer (2012) "Hiring, churn, and the business cycle," *American Economic Review*, 102 (3), 575–79.

Lerner, Josh and Julie Wulf (2007) "Innovation and incentives: Evidence from corporate R&D," *the Review of Economics and Statistics*, 89 (4), 634–644.

Lucas Jr, Robert E and Benjamin Moll (2014) "Knowledge growth and the allocation of time," *Journal of Political Economy*, 122 (1), 1–51.

Marco, Alan C, Joshua D Sarnoff, and Charles deGrazia (2019) "Patent Claims and Patent Scope," *Research Policy*, 48 (9), 103790.

Miranda, Javier and Nikolas Zolas (2017) "Measuring the impact of household innovation using administrative data," in *Measuring and Accounting for Innovation in the 21st Century*: University of Chicago Press.

Molloy, Raven, Christopher L Smith, and Abigail K Wozniak (2014) "Declining migration within the US: The role of the labor market,"Technical report, National Bureau of Economic Research, WP-20065.

Monath, Nicholas, Christina Jones, and Sarvo Madhavan (2021) "Patents View: Disambiguating Patent Inventors, Assignees, and their Locations,"Technical report, American Institute for Research.

Ouimet, Paige and Rebecca Zarutskie (2014) "Who works for startups? The relation between firm age, employee age, and growth," *Journal of financial Economics*, 112 (3), 386–407.

Perla, Jesse and Christopher Tonetti (2014) "Equilibrium imitation and growth," *Journal of Political Economy*, 122 (1), 52–76.

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

Schmookler, Jacob (1957) "Inventors past and present," *The Review of Economics and Statistics*, 39 (3), 321–333.

Toivanen, Otto and Lotta Väänänen (2012) "Returns to inventors," *Review of Economics and Statistics*, 94 (4), 1173–1190.

Tsusaka, Miki, Martin Reeves, Stephanie Hurder, and Johann Harnoss (2017) "Diversity at Work," `https://www.bcg.com/publications/2017/diversity-at-work`.

USPTO (2019) "Progress and Potential: A profile of women inventors on U.S. patents,"Technical report, USPTO.

Vilhuber, Lars (2018) "LEHD Infrastructure S2014 files in the FSRDC,"Technical report, Center for Economic Studies CES-WP-2018-27.

Wagner, Deborah and Mary Lane (2014) "The person identification validation system (PVS): applying the Center for Administrative Records Research and Applications' (CARRA) record linkage software,"Technical report, Center for Economic Studies, US Census Bureau, WP-2014-01.

Winston, Sanford (1937) "Bio-social characteristics of American inventors," *American Sociological Review*, 2 (6), 837–849.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi (2007) "The increasing dominance of teams in production of knowledge," *Science*, 316 (5827), 1036–1039.

Online Appendix for

Measuring the Characteristics and Employment Dynamics of U.S. Inventors

Ufuk Akcigit Nathan Goldschlag

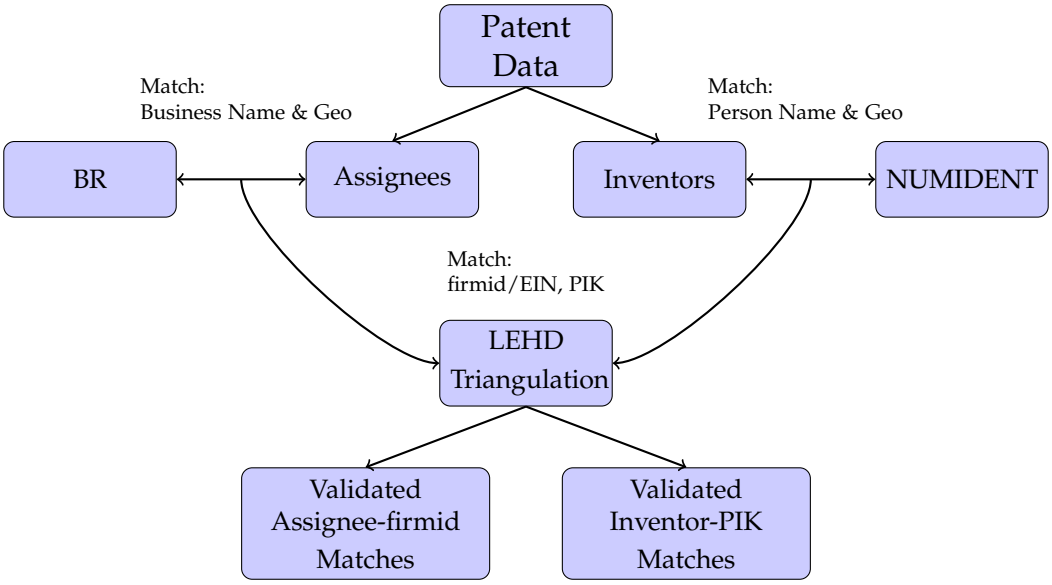
A.1 Triangulation Matching Methodology

The methodology used to match patent assignees and inventors to Census Bureau business and person microdata was first introduced by Graham et al. (2018) (henceforth GGIMM) and later extended by Dreisigmeyer et al. (2018) (hencforth DGKOP). Since DGKOP is a technical note, only available to users with approved projects via the FSRDC system, we summarize the key, non-confidential components of that paper here. One of the primary purposes of DGKOP is to document more precisely, as technical specifications, the steps used to generate the triangulation patent-assignee crosswalk, providing additional details on the processing steps not available in GGIMM. It also contains descriptions of additional match passes that extend the work of GGIMM as well as undisclosed match quality metrics that give users with approved access additional information about the quality of the firm-assignee matches made. In particular, DGKOP describe the additional models B3, FD2, and FA1. We provide a brief summary of each below.

The overall structure of the programs used to produce the patent assignee-firmid and inventor-PIK linkages remains consistent with GGIMM. Figure A1 summarizes the matching process. First, we collect information from the patent documents on assignees and inventors. For patent assignees we collect business name and geography, translating city and state combinations into 3-digit zip codes. For patent inventors we collect person name and geography, again translating city and state into 3-digit zip codes. We then match patent assignee information to the Business Register (BR), the Census Bureau's universe sampling frame of non-farm employer business establishments and the source of information used to create the Longitudinal Business Database (Jarmin and Miranda, 2002; Chow et al., 2021). The patent assignee-BR match relies on fuzzy name and geography matching methods that are relatively permissive, allowing many false positives. This relatively permissive match ensures true positives are among the matched set. We submit the inventor name and geography information to the Person Validation System (PVS), which performs probabilistic matching of person information to the Census Numident reference files (Wagner and Lane, 2014). The Census Numident contains one record for each Social Security Number (SSN) along with all variants of date of birth, name data, and is enhanced with address information from administrative records. The PVS match yields linkages between inventor records and protected identification keys (PIKs). Finally, there is a second match of the patent assignee information to the BR

that limits false positives. This match, which is called the "sensitive" match, is used to validate BR-only matches and augment the crosswalk when it is able to find matches that the triangulation match could not.

FIGURE A1: TRIANGULATION MATCHING FLOW



Both the patent assignee-firm and inventor-PIK linkages that result from the initial BR and PVS matching are noisy. These matches rely on relatively loose criteria and contain more false positives than would be typically desirable in a Census microdata crosswalk. However, these false positives are a feature, not a bug, when combined with the LEHD jobs data. Allowing these false positives increases the probability that true matches are among the set of all matches on both sides (assignee and inventor) when "triangulated" with LEHD jobs information. The next step of the matching collects firmid/EIN-to-PIK linkages from the LEHD earnings history for all PIKs found in the PVS matches. We then query the jobs for those individuals where the firmid/EIN matched to a given patent-assignee record (via the BR match) matches the firmid/EIN-PIK earnings records in the employment histories of the PVS-matched individuals. Matches are identified and flagged with "model types" based upon what information was found to be in agreement between the assignee and inventor information. The model types are as follows.

Models A1-3, Fully Triangulated Matches flagged as A1 occur when the firmid and EIN matches for both the inventor-PIK employment history and the BR-assignee match. A2 matches are cases where the firmid matches but the EIN does not and A3 are cases where the EIN matches but the firmid does not.[24] We also extract unique patent inventor-to-PIK matches from the A1-3 models, which becomes the backbone of our inventor-PIK

[24]The firmid variable on the LEHD is drawn from a match of the LBD to LEHD, as described by Haltiwanger et al. (2016). Approximately 70% of LEHD records are assigned a firmid in this process.

crosswalk. The A models are the only ones that generate inventor-PIK matches. The remaining models are used to generate assignee-firmid matches.

Models B1-3, BR Only Matches are flagged as B1 if there is a BR match but no LEHD match and the matched firmid is among the set of firmids that are fully triangulated (A1-3) within the grant year. We call these BR only A1-3 validated. B2 are cases that do not meet the B1 criteria but for which the same match appears in the "sensitive" BR-patent assignee match. The "sensitive" assignee matches are used to weed out bad matches from the initial pool of B2 matches. Finally, B3 matches are those where a match appears in the "sensitive" match but not in the triangulated match.

Models C1-3, LEHD Only Analogous to B1 models, the C1 cases are those where there is only an LEHD-side match and the matched firmid appears among the fully triangulated A1-3 matches. C2 matches occur when the LEHD only match is validated across multiple inventors—all of the inventors on the patent are also observed working for the LEHD-only firmid match in the patent's application year. C3 cases are the residual LEHD-only matches where a unique firmid is observed.

Models E1-2, Disagree but with Unique Matches E1 matches occur when the assignee-BR and LEHD firmids and EINs disagree but a unique firmid is observed among the BR matches. E2 matches occur when the assignee-BR and LEHD firmids and EINs disagree but a unique firmid is observed among the LEHD matches.

Models D1-2, Name and Clerical For D1 matches, we create an assignee name-to-firmid crosswalk using the high-quality A1-3 matches and then match by assignee name to patents that did not receive a match in the model types A, B, C, and E. D2 matches are made using a clerical database of assignee name-to-firmid matches. Clerical candidates were originally generated by finding the unmatched assignee names associated with the most patents.

Models FD2 & FA1, D2 and A1 Extensions The F models attempt to extend the reach of manual matches (D2s) and the high quality matches (A1s). For FD2, we utilize additional standardizations of the assignee name strings drawn from the USPTO "DVD data" and XML patent data. The "DVD data" contain significant UPSTO manual name disambiguations. These additional versions of the name, when combined with the clerical matches used for the D2 matches, create a name-to-firmid crosswalk. We use that crosswalk to assign firmids to unmatched patents. For FA1 matches we relax the uniqueness constraint placed on the D1 models. Specifically, we find cases where 70% or more an assignee name string's match to to a single firmid (via A1-3 matches). We again use this assignee name-to-firmid crosswalk to match firmids to unmatched patents.

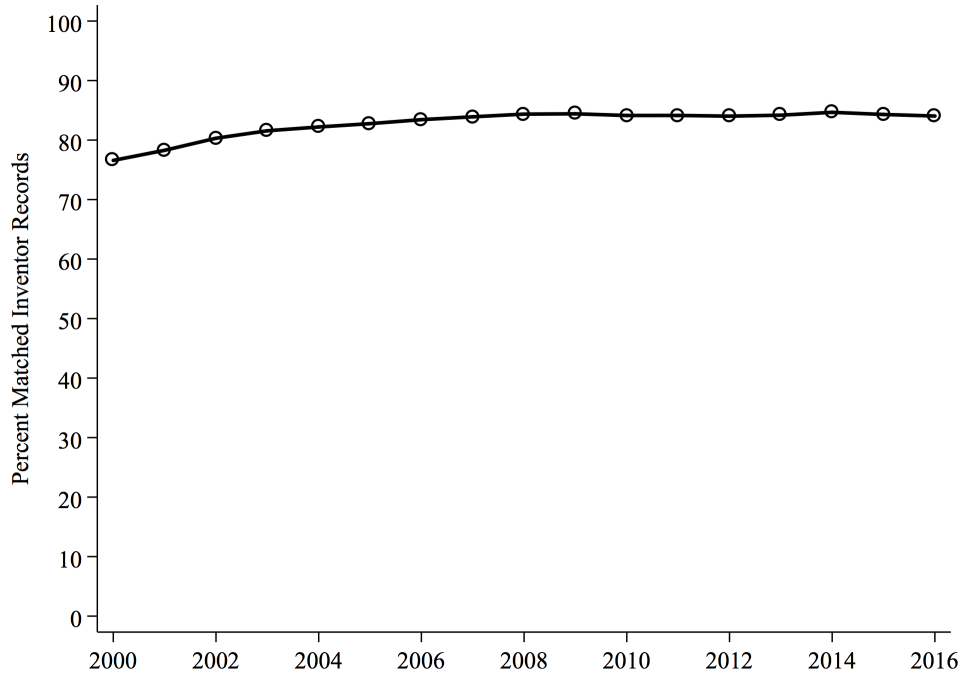
After all match steps have been executed we perform the additional match steps described in Section 3 to combine and clean inventor matches. In particular, we combine A1-3 triangulated inventor-PIK matches with unique PVS matches and utilize PatentsView inventor identifiers to extend the reach of those matches.

A.2 Additional Match Quality Statistics

In this section we provide measures that characterize the types and source of matches and their stability over time. As described in Section 3, if an A1, A2, or A3 match cannot be made for an patent-assignee-inventor combination we still attempt to assign a PIK based on either a unique PVS match or by levering PatentsView inventor identifiers to extend the reach of PIKs we've already matched. A unique PVS match will occur when the inventor's name and geography are sufficiently unique that the PVS system returns a single matched individual. This can occur either because the geography is sufficiently small or the name is sufficiently unique. Census data use and disclosure policies limit the extent to which we can characterize the types of names that do or do not match, but there are several analyses below that provide a sense of match quality.

First, as shown in Figure A2, the match rate is quite stable over time. In 2000 about 77% of US-based inventor records had a matched PIK, which rose to about 84% by 2016. In addition to the patent-level characteristics shown in Table 2, one might worry that matched and unmatched the inventor records differ. One dimension that may affect the match is the size or density of the inventor's geography. On the one hand, higher density may make it more difficult to find a unique PVS match. On the other hand, higher density may be correlated with patent quality metrics, and higher quality patents are more likely to be matched (Table 2). The match rate for inventor records in metropolitan statistical areas (MSAs) is about 84% while the match rate among non-MSAs is about 75%. This 9 p.p. gap suggests that our matches are more likely to be drawn from larger metropolitan areas, but not overwhelmingly so.

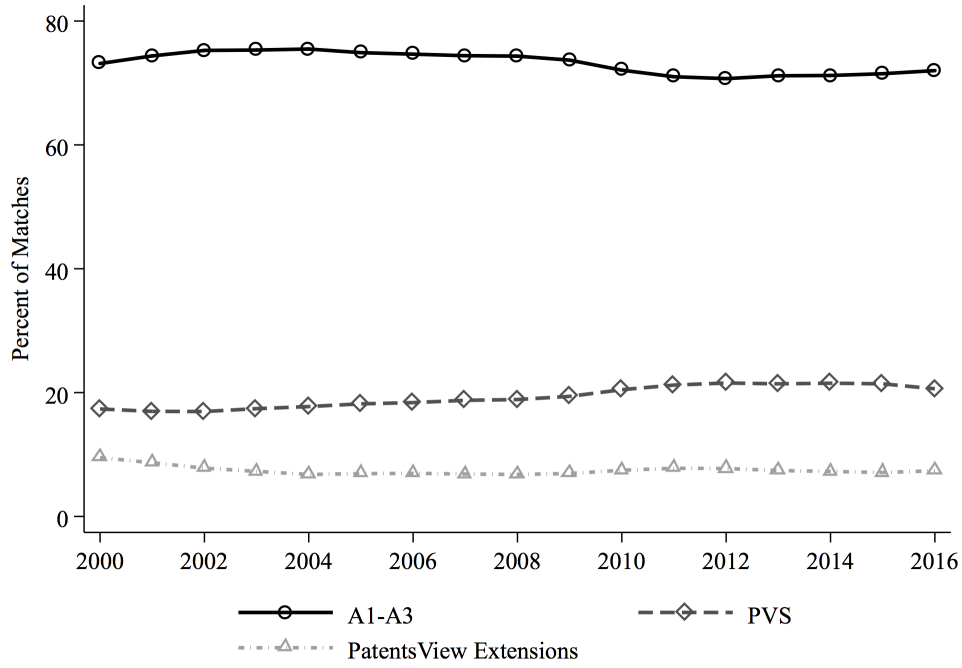
Figure A2: Inventor Record Match Rate by Year



Source: Inventor Record-PIK Match
Notes: Share of inventor records by grant year that receive a PIK match. The match rate in this figure differs slightly from the 93% listed in Section 3 because this figure conditions on records with (1) an assignee match model type and (2) that can be assigned as in metropolitan statistical area or not.

Approximately 19% of matches are assigned using PVS uniques, about 7% are assigned using the PatentsView extensions, and the remainder are the high quality A1-3 matches. Figure A3 shows the percent of matched inventor records by source and grant year. The composition of the match is stable, with the percent of matches drawn from the A1-3 matches ranging from 71% to 75% over time.

Figure A3: Matched Inventor Records by Match Type and Year



Source: Inventor Record-PIK Match
Notes: Share of matched inventor records by grant year and whether the inventor record-PIK match was made using A1-3, a unique PVS match, or the PatentsView extension by grant year. The match composition shown here differs from that in Figure 1 because this figure conditions on records with (1) an assignee match model type and (2) that can be assigned as in metropolitan statistical area or not.

To better understand the types of records that receive a PVS unique or PatentsView extension match, Table A1 shows the percent of PVS unique and PatentsView extension matches by the model source (as described in Appendix Section A.1) for the assignee match. The majority of PatentsView extension matches, and the plurality of PVS unique matches, are made for patents whose assignee received an A-model match. This can occur when the patent has more than one inventor, and at least one of those inventors has a PVS match with employer information that matches the BR-side firmid match. One of the inventor's on that patent may be employed by a different firm or not employed at all during the relevant time window, in which case they do not receive an A1-3 match. Despite this, the inventor may either (1) have a unique match from PVS, or (2) have an inventor ID linked to other patents for which we were able to match a PIK. The second most common assignee model type for the PatentsView and PVS Only matches is the F models. F-models, involve exact name-based matching from derived name-to-firmid crosswalks. For these cases we were unable to make a triangulation match (A) nor a BR or LEHD-only match (B, C), but instead relied on the assignee name string being assocaited with a firmid in some other match pass.

Table A1: Assignee Match Type for PatentsView and PVS Unique Inventor Record Matches

Model	PatentsView	PVS
A	51.2	39.1
B	8.2	8.6
C	6.3	9.5
D	12	11.6
E	7	13.2
F	15.4	18
Total	263,500	705,500

Source: Inventor Record-PIK Match
Notes: Table shows the percent of PatentsView extension and PVS unique matches by the model type for the firmid match of the (first) associated assignee.