Startups, Unicorns, and the Local Supply of Inventors

Benjamin Balsmeier,[a] Lee Fleming,[b,c] Matt Marx,[d,e] Seungryul Ryan Shin[f]

a) *University of Luxembourg*
b) Haas School of Business, UC Berkeley
c) Fung Institute for Engineering Leadership, UC Berkeley
d) Cornell University
e) National Bureau of Economic Research
f) Seoul National University

September 2021

Abstract: We estimate the impact of the inflow of inventors on the formation and success of local venture-backed startups, strengthening causality with a shift-share instrument based on the historic spatial distribution of millions of inventor surnames in the 1940 U.S. Census. Arrival of inventors increases the number of venture-backed startups, but only in same technology fields of the newly-arrived inventors—and at the expense of other fields. Inventor arrivals boost the number of successful startups—including $1B+ "unicorn" exits—while reducing bankruptcies and "fire-sale" acquisitions. The improvement is also driven by a reallocation of venture capital away from investment in low-tech startups, especially unsuccessful ones.

Keywords: Human capital, productivity, high-growth entrepreneurship, inventors

* The authors thank Guan Cheng Li for invaluable research assistance. We also thank participants in the NBER Productivity Seminar and the 2021 RCEA Future of Growth Conference for comments and suggestions. We gratefully acknowledge financial support from The Coleman Fung Institute for Engineering Leadership, the National Science Foundation (1360228), and the Ewing Marion Kauffman Foundation. Errors and omissions are ours.

INTRODUCTION

Entrepreneurship—especially when driven by novel technologies—has been recognized as an essential source of economic growth and improved quality of life since Smith (1776) and Schumpeter (1942). Recent evidence confirms that newly-founded firms are responsible for job creation (Decker et al., 2014; Glaeser, Kerr, & Kerr; 2015), productivity (Gennaioli et al, 2013), and additional innovation (Kortum & Lerner, 2001). Unsurprisingly, policymakers worldwide have sought to spur startup activity, often in hopes of replicating the entrepreneurial dynamics of California's Silicon Valley. That so many efforts have fallen far short (Lerner, 2009) speaks to a lack of understanding and causal evidence for an earlier stage in the chain: if entrepreneurship drives economic growth, what drives entrepreneurship? Further, given that the vast majority of new firms fail (Haltiwanger, Jarmin, & Miranda, 2013)—including 75% of venture-capital backed firms (Hall & Woodward, 2010)—what are the critical inputs for *successful* startups?

The co-occurrence of the words "innovation and entrepreneurship" is ubiquitous in both academic and popular circles. In this paper, we examine whether *productive* entrepreneurship (i.e., successful startups) depends critically on innovation—or, more precisely, on the inventors who are responsible for innovations.[1] To be sure, scholars have long observed that human capital, including technical talent, is an important ingredient in the entrepreneurial recipe. Lerner & Nanda (2020) claim that "[r]egions like Silicon Valley have an abundance of resources for entrepreneurs, [including] excellent engineers…" Jensen & Thursby (2001) likewise argue that scientific inventors need to be fully engaged and motivated for technologies to be successfully commercialized in new firms (see also Zucker et al, 1998; Marx & Hsu, 2021).[2] Larger-scale, if suggestive, evidence for the role of inventors in high-growth entrepreneurship comes from correlations between the supply of technical workers' levels of patenting, entrepreneurial firm founding, and employment (e.g. Kerr, 2013; Maloney & Caicedo, 2016; Azoulay et al., 2020). Glaeser & Kerr (2009) find that talent explains 60-80% of the variance in regional entrepreneurship in U.S. manufacturing, concluding that "the broad stability of this finding suggests that people and their human capital are probably the crucial ingredient for most new entrepreneurs" (p. 659).

Indeed, even absent causal evidence it might seem self-evident that inventors play an essential role in high-growth entrepreneurship. Steve Wozniak, who invented what would become the Apple I while

[1] Our focus on inventors builds on recent advances in disambiguation, which enabled identification of these inventors and facilitated progress on the question of how individuals contribute to innovation, productivity, and economic growth (Bhaskarabhatla et al., 2021; Kline et al. 2019; Azoulay et al. 2020).

[2] Not all high-growth firms in the U.S. are high-tech, and vice-versa. However, Hathaway (2018) reports that high-tech firms are overrepresented by 4x among high-growth firms (21% vs. 5% of all firms) as defined by *Inc*. Magazine's annual list of the 5,000 fastest-growing privately held firms in the U.S (see also Lerner and Nanda, 2020).

working at Hewlett-Packard, famously could not convince his superiors to commercialize the invention and subsequently left to found a new firm with Steve Jobs. At the same time, several successful startups including Slack, Skype, Whatsapp, Alibaba, and BaseCamp, largely contracted out engineering and product development activities to geographically distant locations, and investors regularly pressure their portfolio companies to outsource technical development. As Jim Breyer, managing general partner of Accel Partners, remarked: “There isn't a board meeting that goes by that we don't ask, *Why aren't you being more aggressive [with software development] in India and China?*"[3] Therefore, the direction of causality between the supply of technical talent and entrepreneurship remains unclear (Burchardi et al., 2020). The correlations found by Glaeser & Kerr (2009) could reflect not a causal effect of talent on entrepreneurship but rather the flocking of skilled workers to opportunity. Or, it might be that investors like Jim Breyer are correct and technical talent is simply not as important as conventional wisdom might (like to) assume.

In pursuit of causal evidence on this point, we investigate how the supply of key technical talent—including technology- and task-specific capital (Gibbons & Waldman, 2004)—influences the funding and success of high-growth ventures.[4] We focus on venture-backed startups as a particularly promising subset of new firms. Although only 0.5% of new businesses obtain venture financing (Puri & Zarutskie, 2009), nearly one-half of firms that complete an Initial Public Offering had received venture capital backing (Lerner & Nanda, 2020). We address reverse-causality concerns by instrumenting inventor inflows with the share of inventors’ surnames in a county based on the nationwide distribution of surnames from the 1940 U.S. Census. Our shift-share instrument represents an advance over prior efforts in two ways. First, because the “shares” stem from more than three million unique surnames across more than 3,000 counties, it is less vulnerable to critiques of such instruments with low variation or a few highly-determinative shares (see Goldsmith-Pinkham, Sorkin, & Swift, 2020, Adao, Kolesár, & Morales, 2019, and Borusayak, Hull, & Jaravel, 2018, for a fuller discussion of the issue). Second, focusing on the U.S. lessens concerns regarding endogenous origin-destination combinations (e.g., Indian engineers migrating to Silicon Valley) and also addresses the issue of potential endogenous choice of regions and selection of incoming inventors at the national level (Moser, Voena, & Waldinger, 2014; Parey et al., 2017).

---

[3] https://www.sfgate.com/business/article/looking-offshore-investors-vc-firms-push-for-2813526.php

[4] Related to this paper, several studies have addressed the role of local inventors in regional productivity. For example, Agrawal et al. (2011) show that inventor emigration decreases local knowledge flow in the source region but also drives knowledge back into the departed region. A growing and influential literature on foreign immigration suggests positive impacts on the U.S. of an influx of inventors from outside its borders, including greater patenting and innovation (Bernstein et al., 2018; Hunt & Gauthier-Loiselle, 2010; Burchardi et al., 2020; Kerr & Lincoln, 2010), wages (Peri, Shih, & Sparber, 2015) and TFP (Capelli et al., 2019). Our study differs from these in that we study internal migration and entrepreneurship.

We find that the (exogenous) arrival of inventors in a county has a substantial impact on entrepreneurial activity. Arriving inventors increase the number of venture-backed startups in a county, in the same sectors as the arriving inventors and at the expense of other sectors. Not only does the arrival of inventors produce more startups; the influx of technical talent yields startups with more successful outcomes (IPO or attractive acquisition). Our preferred empirical specifications suggest that counties may expect one additional venture-backed startup for every 28 additional inventors, whereas a successful startup requires an additional 460 inventors. Incoming inventors even contribute to an increase in the number of "unicorn" startups (i.e. exit valuation exceeding $1B). However, the increase in successful exits is not merely the result of more "shots on goal"; these correspond with a reduction in bankruptcies as well as so-called "fire-sale" acquisitions. Therefore, the local availability of technical talent appears to improve the efficiency of venture investment, reallocating away from failed, low-tech startups. These results are robust to a variety of alternative instrument specifications and placebo tests and are moreover not restricted to the top ten counties by entrepreneurial activity (Silicon Valley, etc.).

DATA

We assemble three different sources at varying degrees of aggregation and times to arrive at a panel dataset at the county-year level.

Historic Census data

We begin with the complete 1940 U.S. Census records for 131,940,709 citizens in 38,382,088 households (http://sites.mnhs.org/library/content/1940-census). As we will explain in detail in the next section, our identification strategy relies on being able to observe the name and location of each U.S. citizen in 1940 in order to predict inventor moves. The historic data include 3,363,932 different surnames, of which 27% appear only once. (The median is 3, mean is 39, and maximum is 1,359,079 for Smith.) Figure 1 illustrates the sparse geographical distributions of "Marx". After some cleaning and standardizing procedures, described in detail in Appendix A1, there were 42,268 Flemings, 6,232 Marxes, 153 Shins, and 9 Balsmeiers in the 1940 census data. All analyses below are robust to excluding prolific surnames as indicated by high (local) frequency or wealth, e.g. the Smiths and Rockefellers. The 1940 U.S. Census records consist of 3097 counties and other districts based on the county system in 1940. In order to help matching with the location information of inventors, we translate 19 counties or districts, which are old and no longer in use, to the 2020 concordance. (based on https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2010.html from 1970 to 2020). Please see Appendix A2 for details on geographic disambiguation.

Figure 1 about here

Inventor data

We begin with raw data from the United States Patent and Trademark Office (USPTO) from 1976-2018 (only the intersection of patent and entrepreneurship data are used, see below). Although the USPTO lists inventors for every patent, it does not provide unique identifiers for them. For example, even the relatively rare name of Matthew Marx is listed as inventing many patents, including 5,995,928, "Method and apparatus for continuous spelling speech recognition with early identification, 6,173,266, "System and method for developing interactive speech applications," and 7,271,262, "Pyrrolopyrimidine derivatives." In this simple example, it would seem reasonable based on the titles alone that the same inventor authored the first two but not the last patent, and that is indeed the case. Inventor names can be disambiguated with a variety of algorithms, here we use Balsmeier et al. (2018). After applying the name cleaning and standardizing procedures and the matching algorithm, described in detail in Appendix A1, we match 91.1% of inventors' surname to a surname from the 1940 Census. Note that the name cleaning exercise has no significant effect on the size of the estimated coefficients but decreases matching errors and improves precision of the instrument.

We used the inventor ID and location to identify inventor moves across U.S. counties. We drop all inventors with a single patent. Then, using patent application year as a timestamp, we count an inward move in the first year we first observe an inventor in a county. As noted by Cheyre, Klepper, & Veloso (2015), patent application dates do not necessarily correspond with dates of employment and in particular may lag actual moves. Hence, the inventor may have moved into a county earlier than we detect, leading to a fuzzy lower bound of the actual lag between our variable of interest and the actual inward moves. In 96% of cases, we observe an incoming inventor patenting elsewhere within 5 years earlier (mean = 2.6). Results are robust to excluding inventor moves with longer gaps between two patenting events, or temporary stops at a third county. If an inventor appears on two or more patents within a given year, we follow Moretti & Wilson (2014) and take the most frequent location.

Entrepreneurship data

To measure high-growth entrepreneurship, we use VentureXpert, which is part of Thompson's economic data suite and covers all venture-backed firms in U.S. It offers detailed information on the location, industry classification and significant growth events (M&As and IPOs) of the funded companies. The data is sourced from venture capital firms, company filings and various news sources.

Our baseline sample consists of all startups with available information on founding year, industry and location, starting in 1987 as VentureExpert lacks comprehensive coverage beforehand. Our sample ends in 2007 to avoid truncated measures of whether a startup achieved a significant event (successful M&A or IPO) within ten years since founding. It is worthwhile to note that our sample of venture-backed startups represents a positive selection of startups as VCs typically only fund firms with attractive growth prospects. We focus on such events because they drive economic dynamism, innovation, and long-term economic (Lerner & Nanda, 2020). Although this focus ignores other types of entrepreneurship, e.g. hairdressers, nail polish studios and various sole proprietorships, where productivity growth has been notoriously difficult to achieve because of limited possibilities to leverage technological progress (Baumol & Bowen, 1966), our sample retains low-technology VC backed startups. Figures 2-5 illustrate that the local supply of inventors and high-growth entrepreneurship are indeed strongly correlated, follow similar trends over time, and tend to be regionally clustered.

Figures 2-5 about here

While suggestive, Figures 2-5 cannot speak to whether these patterns reflect self-selection of inventors towards previously successful regions or whether there exists an arguably causal link between the local supply of inventors and entrepreneurship. Furthermore, they fail to differentiate between successful and failed startups and between high-tech (biotechnology, life science, computer and communication and semiconductor industries) and low-tech startups (various categories ranging from food processing to transportation as explicitly defined by VentureXpert). Figures 6 illustrates the spatial distributions of these technological categories and mobile inventors and implies a technology-specific link between inventors and startups that can be exploited both theoretically and econometrically.

Figure 6 about here

Following Ewens & Marx (2018), we define a successful startup as having completed a merger, acquisition, or initial public offering with valuation exceeding 125% of the total invested venture capital within 10 years since founding. We also measure a 500% return on invested capital as well as "unicorns" i.e., startups which exit with a valuation of $1B or greater. Because VentureXpert is missing many acquisition values (and some IPO values), we fill these in using data from Pitchbook and Crunchbase via exact match on website URL and state (Dorn et al., 2020). Where VentureXpert was missing capital invested, we filled in these values from those databases in order to calculate the return on invested capital. We also filled in founding years from the databases when they are missing in VentureXpert.

For failed startups, we used the current status of each startup indicated in VentureXpert, indicating failure if they were listed as "Defunct" or "Bankruptcy." Differentiating between failed and successful startups is crucial, as Decker at al. (2014) show that it is the few high-growth startups that survive the first ten years of their existence that are responsible for about 50% of US gross job creation. Where a startup had not exited within ten years of founding, it was neither counted either as having failed or succeeded.[5]

Table 1 provides descriptive statistics of the dataset at the county-year level. For the sample of 27,619 venture-backed startups, 26% achieve an M&A or IPO within 10 years of their foundation, with an average return of 1646% (median 203%) on the invested capital (note that these values may be over-estimated as we calculated based on only ones that their exit values are available). For failed startups in our sample, there are total 3386 venture-backed startups that are indicated as "Defunct" or "Bankruptcy" in VentureXpert.[6]

Table 1 about here

SHIFT-SHARE INSTRUMENT CONSTRUCTION

We want to find the impact of inventor inflows on entrepreneurship. We can estimate this via OLS:

$$Y_{d,t} = \alpha_0 + \beta \cdot Inv_{d,t-1} + \delta_t \times \eta_d + \gamma_d + \varepsilon_{dt} \tag{1}$$

where $Y_{d,t}$ is a dependent variable observed for county d in year t. $Inv_{d,t-1}$ is the number of inventors who moved to county d in year t-1. δ_t denotes year fixed effects and η_d denotes state fixed effects. We control for state-year specific shocks, such as varying state-level economic conditions and policy changes, through state-year fixed effects $\delta_t \times \eta_d$. γ_d controls for time-invariant unobserved county characteristics that may confound our identification of β. ε_{dt} is the error term.

The key econometric challenge with Equation (1) is that unobserved factors influence both the rate of incoming inventors and local economic conditions; for example, innovative counties are attractive to inventors. Although county fixed effects will effectively control for any persistent differences in innovation levels across counties, this misses temporary local trends that might attract inventors. To address this threat to identification, we construct a shift-share instrument for inventor inflows that builds on the work of Bartik (1991) and its application to international immigration to the U.S. (Card, 2001). Prior studies had noted that immigrants tend to locate near previous immigrants from the same country of

[5] Considering sectors where startups often take longer than 10 years to make it to an exit such as biopharmaceutical industry, we also test with a 12-year window, instead of 10-year window, to capture the successful and failed startups. We use only county-year observations between 1987 and 2005 as we cut the last 2 years to allow for 12 years of observations. The estimation results are almost identical to our main results.

[6] In unreported naïve regressions of local GDP growth on successful and failed startups we find a significant and positive effect of successful startups as opposed to an insignificant effect of failed startups.

origin (Bartel, 1989; Lalonde & Topel, 1991). Card (2001) and others (see Jaeger, Ruist & Stuhler, 2018 for an overview) exploited this observation to predict immigrant inflows into particular regions, by interacting *past shares* of immigrants from an origin country to a given region with the *contemporaneous total inflow or shift* of migrants from the same country at the national level.

We leverage this idea to create an instrument for the contemporaneous inflow of U.S. inventors to a certain county based on the spatial distribution of U.S. surnames across counties in 1940. The intuition is simple: although a host of factors influence where inventors locate—or, more important to our study, *re*-locate—on the margin, an inventor should prefer to move to a county where there are likely to be more relatives. Although we lack data on family structure and relationships for the entire population of U.S. inventors, we borrow an approach from the immigration literature which utilizes the observation that people with a certain family name are found more frequently at places where there were other people with same name in the past (see Darlu, Brunet, & Barbero (2011), for the example of Savoy, France and Clark & Cummins (2015), for England). We illustrate below that these patterns hold for individual US inventors. Specifically, we define our instrument as:

$$\widetilde{Inv_{dt}} = \sum_n \frac{P_{dn}^{1940}}{P_n^{1940}} \cdot Inv_{nt} \qquad (2)$$

where P_{dn}^{1940} is the population of people in county *d* with surname *n* in 1940, P_n^{1940} is the number of people with surname *n* in the entire U.S. in 1940 and Inv_{nt} is the number of inventors with surname *n* who move from any county in the U.S. to any other county in the U.S. in year *t*. The expected inflow of inventors $\widetilde{Inv_{dt}}$ in county *d* at time *t* is thus the weighted sum of inventors that move across the U.S. with surname *n* (the "shift") with the historical distribution of the same family names (the "shares") serving as weights. The intuitive appeal behind this instrument (as in prior immigration studies) is that it generates variation at the local level by exploiting variation at the national level, which is arguably not influenced by local conditions. (That is, the total number of inventors with the name Fleming who move from within the entire U.S. is unlikely to be driven by the local economic conditions of one out of the more than 3,000 U.S. counties.)

Variation and non-persistence of the county-level instrument

One advantage of this instrument over prior shift-share instruments generally, and settlement instruments in particular, is the greater variation in the distribution of names (i.e., the "shares") that stem from more than 3 million unique surnames in 1940 across varying destination and origin areas. (By contrast, immigration studies typically analyze 192 different countries, often with particularly influential origin-destination relationships.) Our estimation should therefore be less vulnerable to problems that arise from low independent variation in shares or overly strong influences of a single or few shares (see critiques in the recent literature, Borusyak, Hull, & Jaravel (2018), Goldsmith-Pinkham, Sorkin, & Swift (2020), or Adao, Kolesár & Morales (2019), and our placebo tests below).

A second advantage of our U.S.-focused shift-share instrument is that a given surname is typically not bound to a specific county of origin (as is more common with country-level analysis, see Moser, Voena, & Waldinger (2014); Parey et al. (2017)). Thus the spatial distribution of the origin of mobile inventors with a surname varies substantially over time (and is the only variation we exploit in our IV). This makes an endogenous origin-destination combination (such as Indian engineers coming into Silicon Valley over long periods of time) highly unlikely to drive our results. Put differently, that mobile inventors with certain names come from various origin counties means that it is less likely that our "shift" is correlated with unobserved endogenous characteristics of origin areas. The considerable variation in the distribution of surnames over time also addresses the "persistence problem" with shift-share instruments in the immigration literature (Jaeger, Ruist, & Stuhler, 2018). Our instrument thus minimizes serial correlation between specific origin and destination regions, as criticized in studies of international migration.

Figures 7-9 illustrate the variation over time and space with the example of all inventors that moved across the U.S. between 1976 to 2015 and have the last name Fleming (75 moves in total). Figure 7 shows how the number of mobile inventors with surname Fleming varies over time yet does not exhibit a trend. The maps in Figure 8 show to which counties the Flemings moved to in the 1980s, 1990s, and 2000s, and the maps in Figure 9 show the origin counties the Flemings moved away from in the 1980s, 1990s, and 2000s, illustrating significant variation in origin and destination counties over time. In fact, to our eyes only one or two of the counties from which Flemings emigrated in the 1990s was also a significant source of Flemings in the 2000s (Figure 9). The same appears true for destination counties in Figure 8, as just one example of why our county-level instrument should be less susceptible than a country-level instrument to the "persistence problem."

Figures 7, 8, and 9 about here

Final instrument with "leave-out"

A remaining concern could be that at least some national movements of inventors are still driven by local economic conditions, and that these might be correlated with past shocks. It could be, for instance, that inventors and families with the name Fleming were always interested in mechanical engineering and thus would have settled in areas where mechanical engineering was in high demand in 1940. If the same area experiences a high demand in mechanical engineering today, then inventors with the name Fleming might be more likely move to that region for endogenous reasons. To reduce these endogeneity concerns, we leave out county d's own inflows from the national flow of inventors with the same surname (see Buchardi et al. (2020), Wozniak & Murray (2012), or Hunt (2017) for similar approaches). Our preferred instrument is thus:

$$Inv_{d,t\widetilde{-1,leave}-out} = \sum_n \frac{P_{dn}^{1940}}{P_n^{1940}} \cdot Inv_{n,t-1,leave-out(n,d)} \tag{3}$$

where $Inv_{nt,leave-out(n,d)}$ is the total number of inventors with name n who move to counties outside of d. The leave-out strategy ensures that the potentially-endogenous choice of Flemings to move to county d does not drive changes in our instrument.

It should be noted that both stages of our IV include county fixed effects. Identification thus derives from weighted time-varying changes in the number of moving inventors for a given surname at the national level, excluding those moving to county d, combined with representation of the same surname in county d in 1940. Ideally, $Inv_{dt,leave-out}$ is truly exogenous and can be used to estimate the causal impact of inventor inflows on using Equation (1), instrumenting $Inv_{d,t}$ with $Inv_{dt,leave-out}$ as in (3).

Industry-specific version of the instrument

Although most of our analyses focus on the overall number of inventors who move into a certain county, we are also interested in the industry-specific inflow of inventors and their influence on the industry-specific rate of startup foundation, e.g. how many biotech startups are founded in a county in response to the inflow of inventors with a biotech background. To this end, we create an additional dataset at the destination county-industry-year level. We differentiate between each of the four high-tech classifications and the low-tech sector as defined by VentureXpert. We match inventors with these industries based on the technology classification assigned to each patent. If an inventor filed patents in more than one tech class we used the most frequent and in case of a tie the earliest (see Appendix A3 for details). Armed with this dataset we can, similar to (1), estimate the following equation with OLS:

$$Y_{d,i,t} = \alpha_0 + \beta \cdot Inv_{d,i,t-1} + \gamma_d \times \delta_t + \theta_i \times \delta_t + \theta_i \times \gamma_d + \varepsilon_{dt} \tag{4}$$

where $Y_{d,i,t}$ stands for a dependent variable observed for county d, industry i at time t. $Inv_{d,i,t}$ is the number of inventors with a technological background closely related to industry i that moved to county d in year t-1. The key difference to (1) is that we can control for county-year specific shocks through county-year fixed effects $\gamma_d \times \delta_t$, i.e. we can effectively control for any unobserved county characteristic, irrespective of whether it is varying or not varying over time. This includes, for instance, the total number of inventors. Put differently, identification of β will only come from relative differences across industries within a county and year. Hence, we only expect β to be positive if, for instance, a higher *fraction* of biotech inventors out of all inventors moving into a given region at a given time would lead to a higher *fraction* of biotech startups within the same region and at the same time. To absorb unobserved industry-specific trends we add industry times year fixed effects $\theta_i \times \delta_t$, and to address unobserved industry-specific advantages or disadvantages of certain places we add industry times county fixed effects $\theta_i \times \gamma_d$. Since all fixed effects enter the first and second stage of our IV regressions, they should further alleviate concerns with respect to unobserved trends in the attractiveness of certain regions that influence the movement of inventors, e.g. Silicon Valley for computer scientists. The match between industry-specific human capital with industry-specific entrepreneurial activity should also reduce measurement error, so we expect β to be larger when estimated with (4) than with (1).

First-stage instrument plausibility check (using individual -level regressions)

Before applying our instrument at the county or county-industry level to obtain results, we first establish the plausibility of its first stage by investigating the linkage between the historical surname distribution and the geographical mobility of individual inventors. This approach rests on an extensive demographic literature, including the migration of people, social networks and mobility (Rossi, 2013). For example, Piazza et al. (1987) tracks migration rates using surname distribution in Italy, Degioanni & Darlu (2001) infer the geographical origin of migrants in a given area using surnames, and Darlu, Brunet, & Barbero (2011) show that surname distribution can be used to estimate mobility using the example of Savoy, France. Studies also use surnames to investigate social mobility, e.g., whether social status changes over centuries (Clark & Cummins, 2014) and whether wealth moves over generations (Clark & Cummins, 2015). In a recent study, Grilli & Allesina (2017) perform a surname analysis on academic professors to compare academic systems in the U.S., France, and Italy.

Our IV approach rests on the assumption that historic surname shares can discriminate between destination counties of moving inventors with a given last name, conditional on moving.[7] We empirically

[7] Adding to the plausibility of our instrument, we also find that the historical share of the same surname in a given location is negatively associated with the inventor's emigration from the location. This supports the argument that inventors are not only more likely to move to regions with a higher historic share of the same surname but also more likely to stay in a region in which more of their families and relatives have resided. Several additional analyses verify the robustness of the results and suggest

test this assumption by estimating a dyadic model that reflects the complete choice set of a moving inventor. To this end, we construct a dataset at the inventor-origin-destination county level that contains each potential destination county combined with the actual county a given inventor is emigrating from. We mark the county the inventor actually moved to with a dummy and for the actual and each potential destination county, include the share of people in the 1940 Census with the same surname. Armed with this dyadic dataset covering 258,657 moves from 1988-2014, we estimate the following model with OLS:

$$Pr(d.cty\#o.cty_{i,o,d,t} = 1|Move\ out_{o,t}) = \alpha_0 + \beta \cdot (\frac{P_{dn}^{1940}}{P_n^{1940}}) + \delta_t + \gamma_d + \gamma_o + \varepsilon_{i,d,o,t} \qquad (5)$$

where $Pr(d.cty\#o.cty_{n,o,d,t} = 1|Move\ out_{o,t})$ is a dummy indicating the destination county ($d.cty$) a given inventor *i* with name *n* moved to from origin county $o.cty$ in year *t*. P_{dn}^{1940} is the population in county *d* with surname *n* in 1940; P_n^{1940} is the population with surname *n* in the entire U.S. in 1940; δ_t denotes a full set of year fixed effects to control for varying macroeconomic conditions; γ_d controls for time-invariant unobserved destination county characteristics; and γ_o controls for time-invariant unobserved origin county characteristics that may confound our identification of β, and $\varepsilon_{i,d,o,t}$ is the error term. We estimate four versions of Equation (5): (a) only with year fixed effects; (b) year and destination-county fixed effects; (c) year and origin-county fixed effects; (d) year and destination-origin county combination fixed effects. Variant (d) absorbs time-invariant county-pair relationship characteristics including, for instance, the geographic distance between two counties. Table 2 presents the results.

Table 2 about here

Although we cannot interpret our LPM specification as a probability model, all specifications consistently show that an increase in the historic surname share in a potential destination county leads to a significantly higher probability of observing a given inventor moving to that specific destination county as compared to all other potential destination choices. The results in Table 2 support the plausibility of our instrument. (Since the dependent variable vector is sparse a low R^2 is to be expected.) The increase in explained variation when destination and destination-origin county fixed effects are included reinforces that unobserved time invariant factors also explain mobility decisions.

---

conditions in which the historical surname effect is moderated. The surname effect is amplified as the average value of houses owned by individuals with the same surname in the county increases, as the foreign-born ratio of individuals with the same surname in the county decreases, or when the inventor resides in a state that enforces non-compete agreements. We find no evidence that the surname effect is susceptible to invention-related inventor characteristics, such as invention productivity, quality, or years of experience as an inventor.

RESULTS

We begin in Table 3 by analyzing the impact of incoming inventors on entrepreneurial *quantity*. These baseline models regress the logged number of venture-backed startups founded in county *d* during year *t* on the logged number of incoming inventors in *t-1* (where t-1 is an upper bound of the actual time of arrival, see above for details). Table 3, model (a) estimates Equation (1) via naïve OLS. Model (b) applies our IV approach with the instrument defined in (3). Model (c) includes state-year fixed effects to absorb unobserved impacts from US states' policy changes and model (d) adds county fixed effects. Model (e) shows estimates that exclude the top 10 entrepreneurial counties including Silicon Valley.[8]

Table 3 about here

Interpreting Table 3, model (a) shows a strong correlation between the number of incoming inventors in a county with the count of venture-backed startups founded the following year, consistent with Glaeser & Kerr (2009). The remaining models (b-d) employ the IV approach and all show a significant positive impact of incoming inventors in a given county on the local rate of startup formation. The strength of the instrument drops somewhat after the inclusion of county fixed-effects; however, the first stage *F* value always remains well above conventional levels, suggesting that the IV regression does not suffer from weak instrument bias (Stock & Yogo, 2002; Lee et al., 2021). In our preferred model (d), the coefficient also drops below the naïve OLS estimate, arguably because the IV reduces bias from self-selection of inventors into more prosperous counties. Model (e) further supports that our results are not limited to Silicon Valley and similar areas. Rather, arriving inventors give rise to more startups generally. Under the assumption that the estimated coefficient can be interpreted as an elasticity, model (d) suggests that a 10% increase in the rate of incoming inventors increases the rate of venture-backed startups founded by 1.8% at the mean. Translating the relative increases into absolute numbers suggests that 10 more inventors lead to 0.035 more startups. Put differently, a county can expect one additional venture-backed startup for every 28.4 incoming inventors.

Is the effect of inventors on firm founding sector specific?

One concern with the baseline analysis is that the linkage between the arrival of inventors in fields unrelated to the industry where startups are founded, which may add measurement error and downward

[8] The top 10 entrepreneurial counties include Alameda County, Los Angeles County, Orange County, San Diego County, San Francisco County, San Mateo County, Santa Clara County in California, Middlesex County in Massachusetts, New York County in New York and King County in Washington.

bias our results. If for example a focal county only had software inventors move in, but all of the increase in startup activity was in biotechnology, we might wonder whether our model accurately enough resembles the notion of an application of task-specific human capital (Gibbons & Waldman, 2004) to relevant new ventures. To this end, we turn to the county-industry level instrument, as described above and formally shown in Equation 4, where inventors are mapped to specific VentureXpert industry categories based on the corresponding technology classes of their patents (Appendix Table A3). The analysis resembles that of Table 3, but the dependent variable is the number of startups (models a and b) founded in industry *i* at a given county *c* and time *t*. The finer unit of measurement leads to an increase in the number of observations although the underlying data source stays the same. Econometrically it has the advantage of allowing absorption of any unobserved shocks at the county level, whether time variant or invariant, through a richer set of fixed effects: county-industry, county-year, and industry-year.

Table 4, model (a), estimates a positive effect on the founding of ventures in the same industry as the inventors of supporting technologies. Table 4 model (a) implies that a 10% increase in the rate of incoming inventors increases the rate of venture-backed startup formations in their field by 5.1% at the mean. The larger positive coefficient, relative to Table 3 model (d), is consistent with a reduction in measurement error. This result suggests that the findings in Table 3 are not spurious due to a generally "rising tide" of startups due to an overall increase in population or supply of technical talent overall; rather, startups arise in the same sectors in which talent has recently been boosted. This supports the inference that an increase in the local supply of technical human capital is causally responsible for entrepreneurial activity in that same sector. It is reminiscent of Bell et al.'s (2020) finding that children are not only more likely to become inventors when they are born in the vicinity of more inventors, but they are more likely to become inventors in the same fields as the inventors they are exposed to.

The field-specific nature of this exposure is further reinforced by model (b), which reveals a negative effect for unrelated technical sectors. This offsetting result makes sense in the context of venture-backed startups, as venture investors must decide how to allocate their dollars. If biotech inventors arrive in the county and biotech startups get funded, it follows that fewer (local) dollars are available for non-biotech startups, as we see in model (b). These results support Lerner & Nanda's (2020) arguments that VCs look for, "…a very narrow band of technological innovations…" (p. 238) and that venture capital reaches a relatively small proportion of entrepreneurial startups.

Table 4 about here

Startup success vs. failure

So far, we have established that the arrival of inventors is responsible for the founding of new firms. Although many governments adopt the number of startups as an easy-to-count metric (Lerner, 2009), to truly contribute to jobs, productivity, and growth one would want to measure *successful* startups. Haltiwanger, Jarmin, & Miranda (2013) note that although startups create many jobs, they also destroy many jobs because failure is the modal outcome. But "success" is not easily discerned. Although Initial Public Offerings almost always indicate a successful startup, acquisitions can be an ambiguous indicator of success. Puri & Zarutski (2012) report that many venture-backed failures are "disguised" as acquisitions, often sold for pennies on the dollar. As noted above, VentureXpert was missing many exit values, so we merged Pitchbook and CrunchBase data with VentureXpert to augment coverage.

In Table 5 we only consider the venture-backed startups founded in county *d* during year *t* as the dependent variable that become successful within a ten-year window. In model (a), "Successful" is determined retrospectively as the number of firms founded that achieved an IPO or were acquired with a 125% rate of return (as per Ewens & Marx, 2018). The estimates from model (a) suggest that a 10% increase in the rate of incoming inventors increases the rate of successful venture-backed startups founded by 1.0% at the mean. Translating the relative increases into absolute numbers suggests that 10 more inventors lead to 0.022 more successful startups. Put differently, a county can expect one additional successful venture-backed startup for every 460 incoming inventors.

The result in model (a) indicates that incoming inventors are not only responsible for an increase in entrepreneurial activity, as in Table 4, but also an upshot in *successful* startups and assumedly the accompanying jobs, innovations, growth, and liquidity events. One might wonder whether these inventors are only responsible for startups that "just barely" succeeded in returning capital to investors, as opposed to generating some of the more spectacular returns and success stories. We further raised the threshold of an exit value to 500% of total venture capital acquired in model (b), which substantially reduces the magnitude of the estimated coefficient but remains statistically significant. In model (c), we show that inventors even give rise to so-called "unicorn" startups with exit values in excess of 1 billion dollars.

Of course, this increase in the number of successful startups—at all levels—could be a mechanical result of "more shots on goal" so to speak. That is, investors place more bets on more startups and win more often. Therefore, we also test how the influx of inventors affects the failure rate of startups, i.e., venture-backed startups founded in county *d* during year *t* that eventually *failed*. In model (d), we use the traditional measure of "failed" startups as those that are currently Defunct or Bankrupt as indicated in

VentureXpert. The results suggest arriving inventors reduce formation of failed startups in the county. Mindful of the Puri & Zarutskie (2012) discovery of failed venture-backed startups "disguised" as acquisitions, in model (e) we include with bankruptcies exits with a valuation lower than 125% of total venture capital invested. Model (e) likewise shows a negative effect of incoming inventors on failed startup foundations (and is robust to eliminating exits with >100% return on investment, or >50%). We conclude that inventors not only causally improve the quantity but also the quality of entrepreneurship.

Table 5 about here

Reallocation from low-tech into high-tech sectors

In Table 6 we dig deeper into the dynamics underlying the reallocation in Table 5 from lower to higher quality investments. In exploring these mechanisms, we are mindful of past findings that venture investors are local in their investment ability (Sorenson & Stuart, 2001), sensitive even to the availability of direct vs. connecting flights (Bernstein, Giround, & Townsend, 2016). Therefore, state- and even county-level investment decisions may be influenced by the local supply of inventors. We separate high-tech (biotechnology, life science, computer and communication and semiconductor) from low-tech ventures as defined by VentureXpert. Models (a) and (d), which resemble Table 3 in using count of startups as the dependent variable, show a shift from low-tech to high-tech startups upon inventor arrival.

Models (b, c and e, f) of Table 6 explore the dynamics of this reallocation from low- to high-tech, breaking down high- and low-tech into Successful vs. Unsuccessful as in models (a) and (d) of Table 5. Model (c) of Table 6 shows a clear shift away from failed low-tech startups. Model (b) shows that successful low-tech startups also decrease in response to arrival of inventors, though the estimated coefficient is much smaller in magnitude than that of failed low-tech startups and also less precisely estimated. This suggests that the shift is primarily away from the failed startups in low-tech industries; in other words, investors appear savvy enough to keep investing in low-tech firms that prove successful, but they avoid less promising low-tech vehicles when inventors arrive. Models (e) and (f) largely echo the results of Table 5, again suggesting that the influx of inventors improves the efficiency of venture investment, reallocating away from failed, low-tech startups toward successful, high-tech startups.

Table 6 about here

Robustness - Alternative instrument constructions

Although the validity of shift-share instruments does not require exogeneity of the shares, and concerns should be lessened by the inclusion of county fixed effects, we nonetheless estimate robustness checks that should further alleviate concerns of potentially-endogenous share characteristics. We re-estimate model (d) of Table 4, replacing the instrument with alternative calculations of the historic name shares (still applying the leave out strategy). Table 7 shows the results for these alternative instruments.

For the first alternative instrument (model a), we consider only people in the 1940 Census that lived in a given county before 1935. We thus effectively enlarge the gap between the shares and the actual moves of inventors and reduce potential correlation between historic and current inventor migration shocks. In model b, we exclude the 50 surnames that appear most frequently in the historic data, which should reduce concerns that correlated shares of two counties may lead to an over-rejection problem (as shown by Adao et al., 2019). In our third construction (model c), we exclude wealthy families of each county as inventors may benefit even generations later from their ancestors' wealth. Using the historic house value in the 1940 Census, we excluded families holding more than 1% of the total house value of a given county.

Our fourth construction (model d) departs from the shift-share approach, instead calculating the inventor's separation from their surname's historic geographic centroid. We use the inverse geographic distance between each county centroid and the geographic centroid for an inventor's surname as weights when constructing the instrument. The distance between a county's centroid and a surname's historic geographic centroid has the advantage of a very low correlation with any future county or inventor specific characteristics. A limitation of this fourth instrument construction is that most surnames are clustered in multiple geographic and typically urban regions. Thus, even if there is one largest centroid, we will calculate distance from it even if a somewhat smaller but much-closer aggregation exists. The shift-share instrument does not suffer from this limitation and remains our preferred instrument.

Table 7 about here

The coefficient sizes remain robust across different specifications, although the strength of the instrument declines in model (d) compared to our original instrument. Especially with respect to our centroid-distance instrument, this is not surprising. That the instrument strength and coefficient size does not decline greatly when excluding particularly influential families supports the assumption that either 1) there is no direct link between the historic name shares and the second stage regression, or 2) the county fixed effects effectively absorb such potentially worrying relationships.

Placebo tests: random reassignment of instrument

Given the relative strength of the instrument, one might wonder whether our IV effectively absorbs unobserved local characteristics and hence leads to an overly strong rejection of the null hypothesis. To address these concerns, we run three placebo tests in the spirit of Adao, Kolesár, & Morales (2019, henceforth AKM). We randomly reassign the instrument in three ways: (1) across the entire sample, (2) across counties within a given year, and (3) across time within a given county. Then, we re-run our baseline model with each placebo 1000 times. Table 8 summarizes the results of the first and second stages. All three placebos consistently show that a random assignment effectively eliminates a significant prediction of incoming inventors in the first stage, and false identification of a causal impact of incoming inventors on the number of successful venture-backed startups in the second stage. Hence, our IV estimates do not seem to suffer from the artificial over-rejection of the null hypothesis as identified in many other applications of shift-share instruments by AKM. The reason would seem to lie in the effective absorption of unobserved time-invariant heterogeneity at the county level.

Table 8 about here

CONCLUSION

We have provided arguably causal evidence regarding how the arrival of inventors influences both the quantity and quality of entrepreneurship. Our shift-share instrument, based on the county-level distribution of surnames in the 1940 U.S. Census, addresses limitations of similar instruments in the international-migration literature. We are able to show a sector-specific uptick in entrepreneurial activity and also tie the arrival of inventors to a rise in successful startups as well as a lowering of unsuccessful startups. Our estimates indicate that approximately 460 new inventors in a county can create a successful startup, and even "unicorn" startups with >$1B exits can be traced to inventor arrivals. The approach further illustrated how venture capital firms shifted their investment towards high technology opportunities, at the expense of unsuccessful low technology opportunities. The shift away from unsuccessful low tech to high tech firm starts held across all U.S. counties—not just Silicon Valley and similar hotspots—as well as a variety of instruments, and measures.

Although this work sought to explain how the supply of inventors influenced high-growth entrepreneurship, it can also speak to the classic question of why industries cluster geographically (Rosenthal & Strange, 2004; Overman & Puga, 2010; Ellison, Glaeser, & Kerr 2010). Much work has validated the Marshallian agglomeration arguments of production economies, labor pooling, and knowledge spillovers, yet that work has often struggled to isolate and estimate causal mechanisms (Glaeser & Kerr 2009). The shift share instrument developed here enabled investigation of one arguably causal linkage; inventor arrival fuels an increase and funding in startups in those inventors' specific industries. Furthermore, if inventors move towards incipient clusters (e.g., semiconductors in Silicon Valley in the 1960s), their impact on field-specific entrepreneurship and venture capital investment could create a feedback dynamic that directly and dramatically fuels industry concentration.

Given the increasing importance of technology, innovation, and the growth of the knowledge economy, these results also imply an ever-increasing role for STEM labor pooling amongst the three classic Marshallian mechanisms. Assuming that inventor immigration to a region bolsters this role, these results would imply that pooling drives investment which could in turn result in the co-location of production assets. Given that knowledge spillovers are localized and probably reliant upon personal inventor communication (Saxenian, 1996; Thompson and Fox-Kean 2005), then inventor pooling should also increase knowledge spillovers. Future research should seek to disentangle the Marshallian mechanisms that drive agglomeration, estimate their feedback effects, and quantify their relative importance.

The mutual reinforcement of these agglomeration mechanisms could partially explain the rapid emergence of Silicon Valley and the growth in inequality across regions in the U.S. (Glaeser & Hausman, 2020; Lerner & Nanda, 2020). Moretti (2012) labels this phenomenon the "Great Divergence" and provides an example of two relatively similar California towns in 1969 – Menlo Park and Visalia. Surprisingly, given their wide differences now across wealth, crime, education, and health measures, the towns had relatively similar incomes and educational levels in 1969. The venture capital firm Kleiner-Perkins founded their operations in Menlo Park in 1972 and became prominent after a series of high-profile successes, including Amazon, Google, and Genentech. Their private success and similar successes by other nearby investors created a striking concentration of wealth (Lerner & Nanda, 2020), for example, for many years, real estate on Menlo Park's Sand Hill road was the most expensive in the world.

Independent from its implications for regional inequality, this work enables a crude estimate of the "value" of an inventor; geography and mobility in this respect simply provide an instrument to get at that estimate. This estimate is obviously sensitive to the region in which it is derived; the value of an inventor surely varies across regions, based on the inventor, the region, and the interaction of the two. Although this work used arrival in a county to back out the value of an arriving inventor, a home-grown inventor might be just as useful to local entrepreneurship (for example, Steve Wozniak already lived and worked in Silicon Valley before founding Apple). Indeed, if a home-grown inventor had easier access to existing networks of friends, family, investors, and fellow entrepreneurs, they might be even more effective at supporting the success of high-tech firms. It would be interesting to explore whether inventor arrival crowds out–or complements—locally grown inventors and entrepreneurship (Azoulay et. al. 2021).

Future work could also estimate how the loss of inventors impacts the source region. Our back of the envelope calculations implied that inventor arrivals enable 17.9% of high-tech entrepreneurship, however, this calculation ignores the probably negative impact on the home regions of the arrivals. Although beyond the scope of this work, a full accounting of these effects might enable an estimate of the social welfare of inventor mobility. This could then inform policy, for example, should policies encourage industries and technologies to cluster, because such clustering improves innovative efficiency (for one example, through increases in knowledge spillovers), or should policies encourage industries to disperse, and hence distribute jobs and wealth in a more geographically equitable way?

Beyond inventor pooling and investment, regional entrepreneurship ecosystems also depend on physical and institutional infrastructure, lawyers, and non-technical entrepreneurial talent. There are surely declining marginal returns as the supply of tech talent outstrips complementary resources needed for entrepreneurship. This can be seen in the negative effects inventors in one field have on the financing of startups in other fields. Future work should investigate whether the arrival of an inventor in one county decreases entrepreneurship in nearby counties, possibly due to competition for complementary resources.

REFERENCES

Adao, R., Kolesár, M., & Morales, E. (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, 134(4), 1949-2010.

Agrawal, A., Kapur, D., McHale, J., & Oettl, A. (2011). Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, 69(1), 43-55.

Azoulay, P., Jones, B. F., Kim, J. D., & Miranda, J. (2020). Age and high-growth entrepreneurship. *American Economic Review: Insights*, 2(1), 65-82.

Azoulay, P., Jones, B., Kim, J. D., & Miranda, J. (2021). Immigration and Entrepreneurship in the United States. *American Economic Review*, Forthcoming.

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., ... & Fleming, L. (2018). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *Journal of Economics & Management Strategy*, 27(3), 535-553.

Bartel, A. P. (1989). Where do the new US immigrants live?. *Journal of Labor Economics*, 7(4), 371-391.

Bartik, T. J. (1991). Who benefits from state and local economic development policies?.

Baumol, William J. and Bowen, William G. (1966) *Performing Arts: The Economic Dilemma*. MIT Press, Cambridge.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2019). Who becomes an inventor in America? The importance of exposure to innovation. *The Quarterly Journal of Economics*, 134(2), 647-713.

Bernstein, S., Diamond, R., McQuade, T., & Pousada, B. (2018). The contribution of high-skilled immigrants to innovation in the United States. Stanford Graduate School of Business Working Paper, 3748, 202019-20.

Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The impact of venture capital monitoring. The *Journal of Finance*, 71(4), 1591-1622.

Bhaskarabhatla, A., Cabral, L., Hegde, D., & Peeters, T. (2021). Are inventors or firms the engines of innovation?. *Management Science*, 67(6), 3899-3920.

Borusyak, K., Hull, P., & Jaravel, X. (2018). *Quasi-experimental shift-share research designs* (No. w24997). National Bureau of Economic Research.

Burchardi, K. B., Chaney, T., Hassan, T. A., Tarquinio, L., & Terry, S. J. (2020). *Immigration, innovation, and growth* (No. w27075). National Bureau of Economic Research.

Cappelli, R., Czarnitzki, D., Doherr, T., & Montobbio, F. (2019). Inventor mobility and productivity in Italian regions. *Regional Studies*, 53(1), 43-54.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1), 22-64.

Clark, G., & Cummins, N. (2014). Surnames and social mobility in England, 1170–2012. *Human Nature*, 25(4), 517-537.

Clark, G., & Cummins, N. (2015). Intergenerational wealth mobility in England, 1858–2012: surnames and social mobility. *The Economic Journal*, 125(582), 61-85.

Cheyre, C., Klepper, S., & Veloso, F. (2015). Spinoffs and the mobility of US merchant semiconductor inventors. *Management Science*, 61(3), 487-506.

Darlu, P., Brunet, G., & Barbero, D. (2011). Spatial and temporal analyses of surname distributions to estimate mobility and changes in historical demography: the example of Savoy (France) from the eighteenth to the twentieth century. In Navigating time and space in population studies (pp. 99-113). Springer, Dordrecht.

Decker, R., Haltiwanger, J., Jarmin, R., & Miranda, J. (2014). The role of entrepreneurship in US job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3), 3-24.

Degioanni, A., & Darlu, P. (2001). A Bayesian approach to infer geographical origins of migrants through surnames. *Annals of Human biology*, 28(5), 537-545.

Dorn, D., Hanson, G. H., Pisano, G., & Shu, P. (2020). Foreign competition and domestic innovation: Evidence from US patents. *American Economic Review: Insights*, 2(3), 357-74.

Ewens, M., & Marx, M. (2018). Founder replacement and startup performance. *The Review of Financial Studies*, 31(4), 1532-1565.

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3), 1195-1213.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). Human capital and regional development. *The Quarterly Journal of Economics*, 128(1), 105-164.

Gibbons, R., & Waldman, M. (2004). Task-specific human capital. *American Economic Review*, 94(2), 203-207.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, 97(2), 498-520.

Glaeser, E. L., & Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain?. *Journal of Economics & Management Strategy*, 18(3), 623-663.

Glaeser, E. L., & Hausman, N. (2020). The spatial mismatch between innovation and joblessness. *Innovation Policy and the Economy*, 20(1), 233-299.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8), 2586-2624.

Grilli, J., & Allesina, S. (2017). Last name analysis of mobility, gender imbalance, and nepotism across academic systems. *Proceedings of the National Academy of Sciences*, 114(29), 7600-7605.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological tools* (No. w8498). National Bureau of Economic Research.

Hall, R. E., & Woodward, S. E. (2010). The burden of the nondiversifiable risk of entrepreneurship. *American Economic Review*, 100(3), 1163-94.

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who creates jobs? Small versus large versus young. *Review of Economics and Statistics*, 95(2), 347-361.

Hathaway, I. (2018). High-growth firms and cities in the US: an Analysis of the Inc. 5000. *Brookings Institution report*.

Hunt, J. (2017). The impact of immigration on the educational attainment of natives. *Journal of Human Resources*, 52(4), 1060-1118.

Hunt, J., & Gauthier-Loiselle, M. (2010). How much does immigration boost innovation?. *American Economic Journal: Macroeconomics*, 2(2), 31-56.

Jaeger, D. A., Ruist, J., & Stuhler, J. (2018). *Shift-share instruments and the impact of immigration* (No. w24285). National Bureau of Economic Research.

Jensen, R., & Thursby, M. (2001). Proofs and prototypes for sale: The licensing of university inventions. *American Economic Review*, 91(1), 240-259.

Kerr, W. R. (2013). *US high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence* (No. w19377). National Bureau of Economic Research.

Kerr, W. R., & Lincoln, W. F. (2010). The supply side of innovation: H-1B visa reforms and US ethnic invention. *Journal of Labor Economics*, 28(3), 473-508.

Kline, P., Petkova, N., Williams, H., & Zidar, O. (2019). Who profits from patents? rent-sharing at innovative firms. *The Quarterly Journal of Economics*, 134(3), 1343-1404.

Kortum, S., & Lerner, J. (2001). *Does venture capital spur innovation?*. Emerald Group Publishing.

LaLonde, R.J. and Topel, R.H. (1991). *Labor market adjustments to increased immigration*. In

Immigration, trade, and the labor market (pp. 167-199). University of Chicago Press.

Lerner, J. (2009). *Boulevard of broken dreams: why public efforts to boost entrepreneurship and venture capital have failed and what to do about it*. Princeton University Press.

Lerner, J., & Nanda, R. (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives*, 34(3), 237-61.

Lee, D. S., McCrary, J., Moreira, M. J., & Porter, J. R. (2021). *Valid t-ratio Inference for IV* (No. w29124). National Bureau of Economic Research.

Maloney, W. F., & Caicedo, F. V. (2016). The persistence of (subnational) fortune. *The Economic Journal*, 126(598), 2363-2401.

Marx, M., & Hsu, D. H. (2021). Revisiting the Entrepreneurial Commercialization of Academic Science: Evidence from "Twin" Discoveries. *Management Science*, Forthcoming.

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.

Moretti, E., & Wilson, D. J. (2014). State incentives for innovation, star scientists and jobs: Evidence from biotech. *Journal of Urban Economics*, 79, 20-38.

Moser, P., Voena, A., & Waldinger, F. (2014). German Jewish émigrés and US invention. *American Economic Review*, 104(10), 3222-55.

Overman, H. G., & Puga, D. (2010). *Labor pooling as a source of agglomeration: An empirical investigation*. In Agglomeration economics (pp. 133-150), edited by Edward Glaeser. University of Chicago Press.

Parey, M., Ruhose, J., Waldinger, F., & Netz, N. (2017). The selection of high-skilled emigrants. *Review of Economics and Statistics*, 99(5), 776-792.

Peri, G., Shih, K., & Sparber, C. (2015). STEM workers, H-1B visas, and productivity in US cities. *Journal of Labor Economics*, 33(S1), S225-S255.

Piazza, A., Rendine, S., Zei, G., Moroni, A., & Cavalli-Sforza, L. L. (1987). Migration rates of human populations from surname distributions. *Nature*, 329(6141), 714-716.

Puri, M., & Zarutskie, R. (2012). On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms. *The Journal of Finance*, 67(6), 2247-2293.

Rosenthal, S.S. and Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2119-2171). Elsevier.

Rossi, A. (2013). Does economic upgrading lead to social upgrading in global production networks? Evidence from Morocco. *World Development*, 46, 223-233.

Saxenian, A., 1996. *Regional advantage*. Harvard University Press.

Schumpeter, J. (1942) *Capitalism, socialism and democracy*. Vol. 36, Harper & Row, New York.

Smith, A. (1776). *Wealth of nations.*

Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology*, 106(6), 1546-1588.

Stock, J. H., & Yogo, M. (2002). *Testing for weak instruments in linear IV regression*. (No. w0284). National Bureau of Economic Research.

Thompson, P., & Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, 95(1), 450-460.

Wozniak, A., & Murray, T. J. (2012). Timing is everything: Short-run population impacts of immigration in US cities. *Journal of Urban Economics*, 72(1), 60-78.

Zucker, L., Darby, M., & Brewer, M. (1998). Intellectual human capital and the birth of US biotechnology enterprises. *American Economic Review*, 88(1), 290-306.

FIGURES

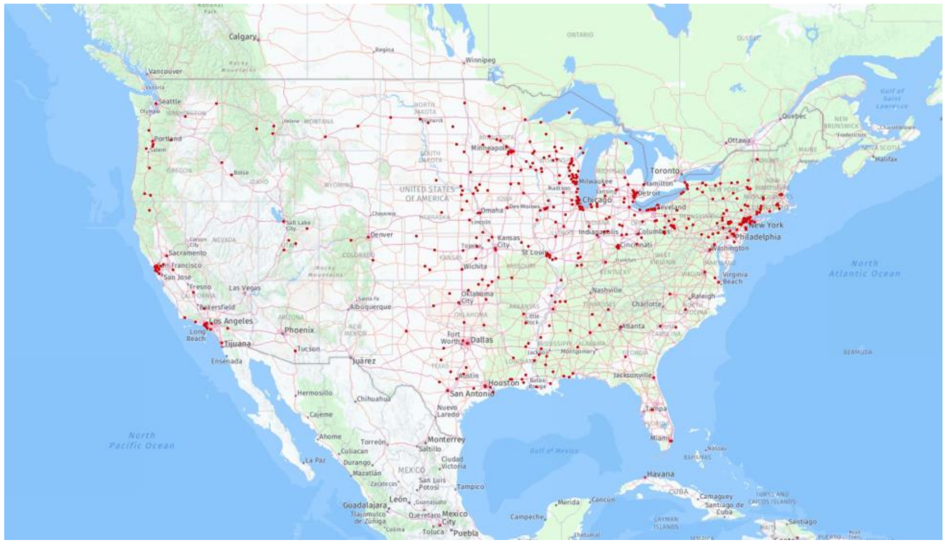

Figure 1: Spatial distribution of the surname "Marx" in 1940 (each red dot = 50 individuals).

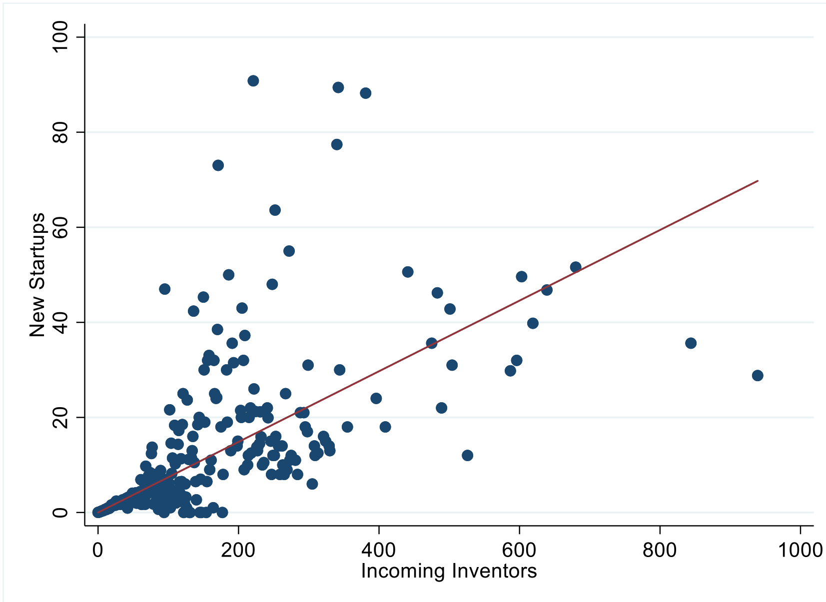

Figure 2 – Graphical representation of raw data

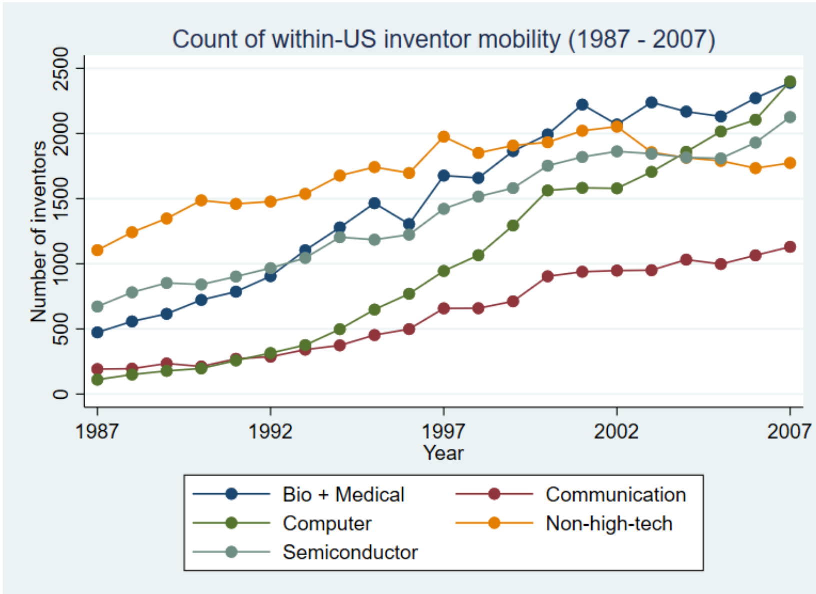

Figure 3 – Graphical representation of US inventor moves

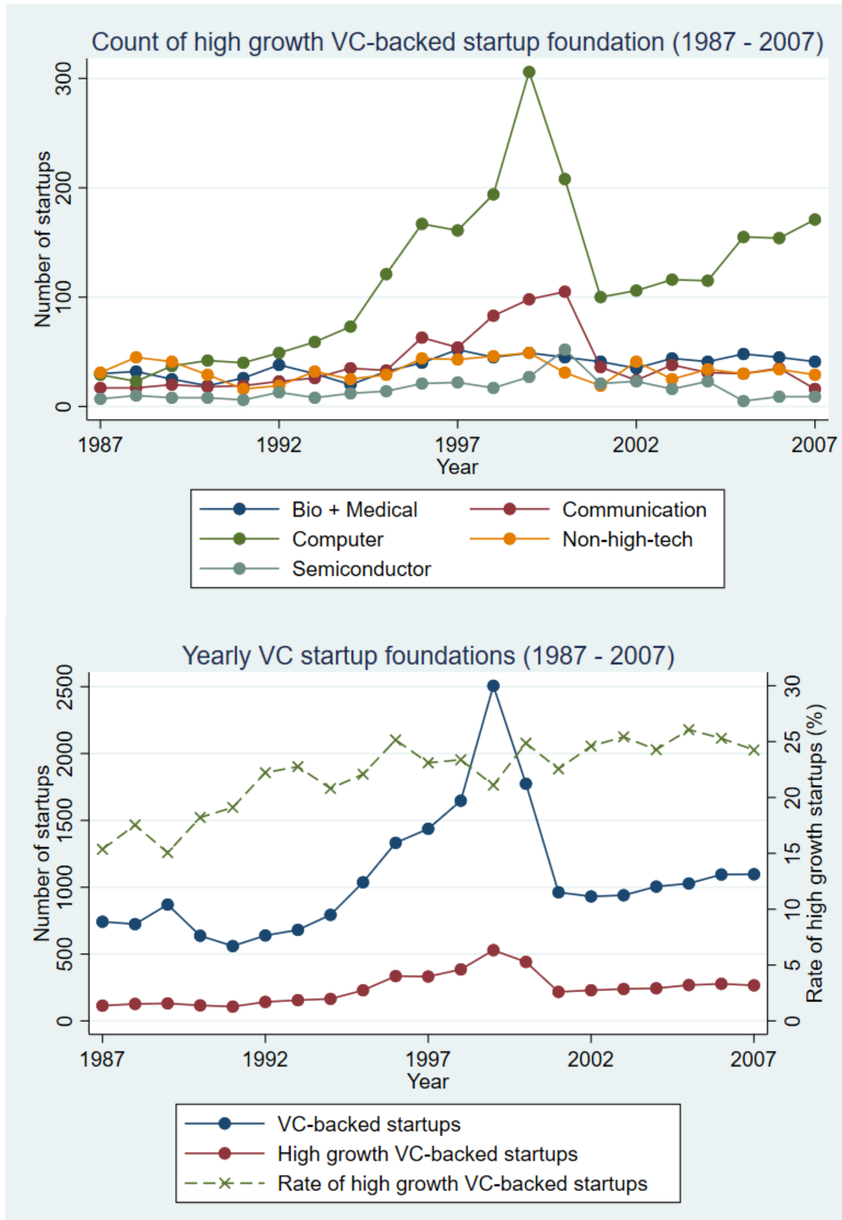

Figure 4 – Graphical representation of venture-backed startup creation

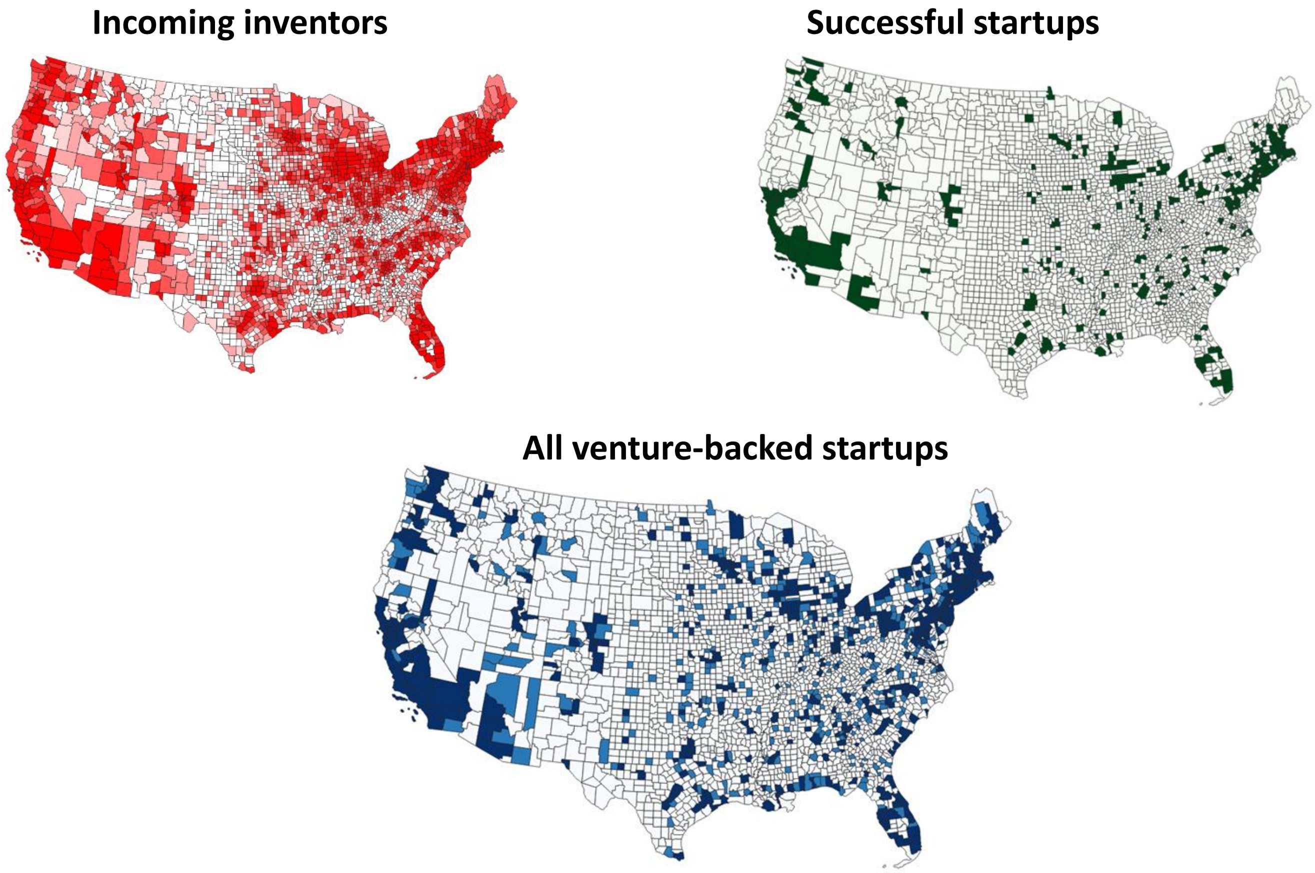

Figure 5: Geographical clustering of inventor moves, startups, and high-growth startups, 1987 to 2007

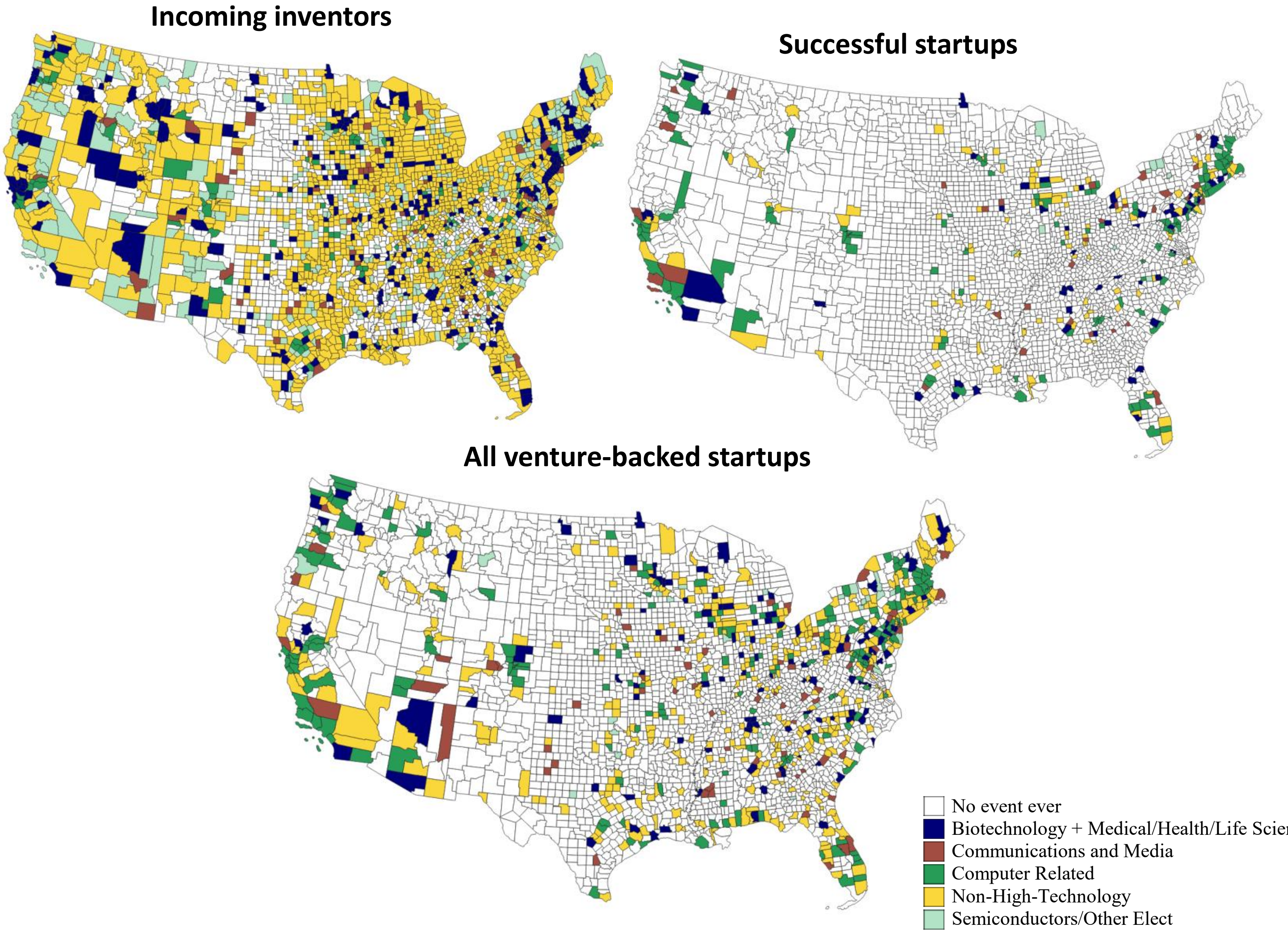

Figure 6: Geographical clustering of inventor moves, startups, and high-growth startups, 1987 to 2007 by major technology within county

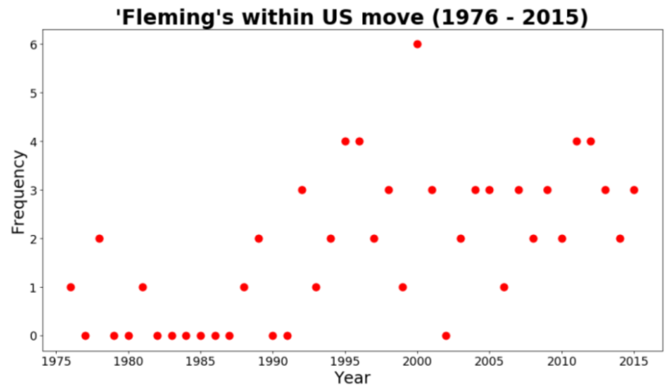

Figure 7: Frequency of moving inventors within the U.S. named Fleming over time

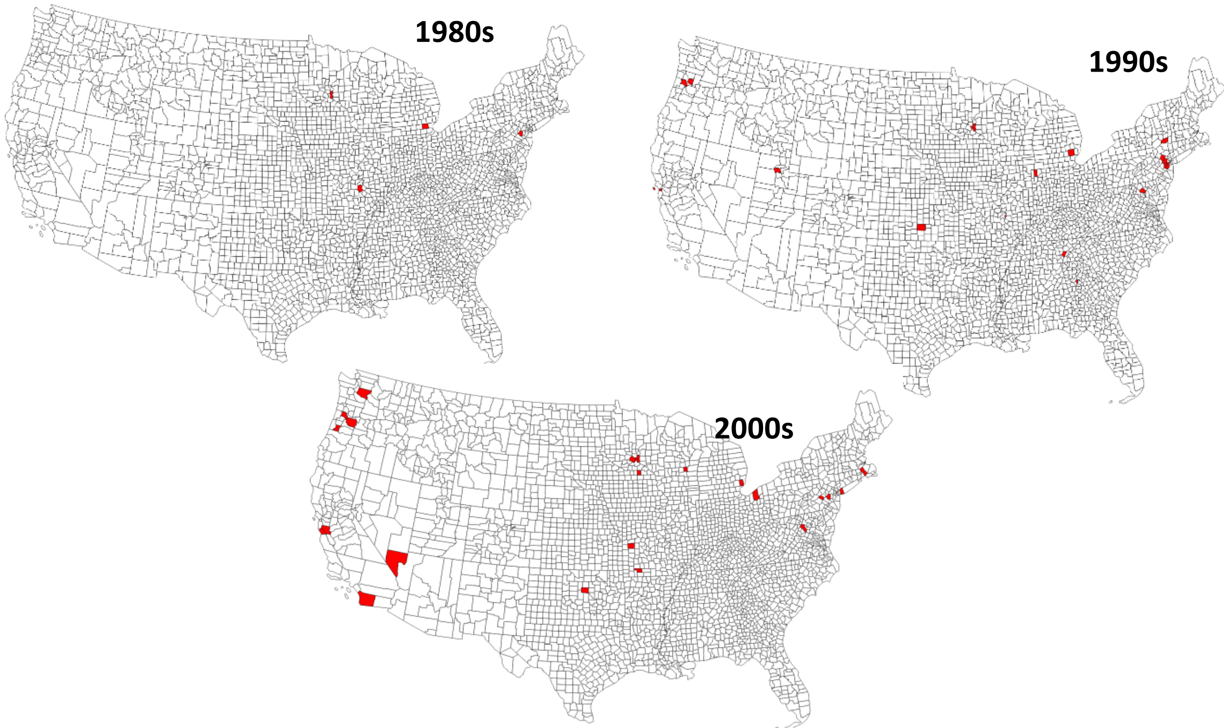

Figure 8: Destination counties of moving inventors within the U.S. named Fleming

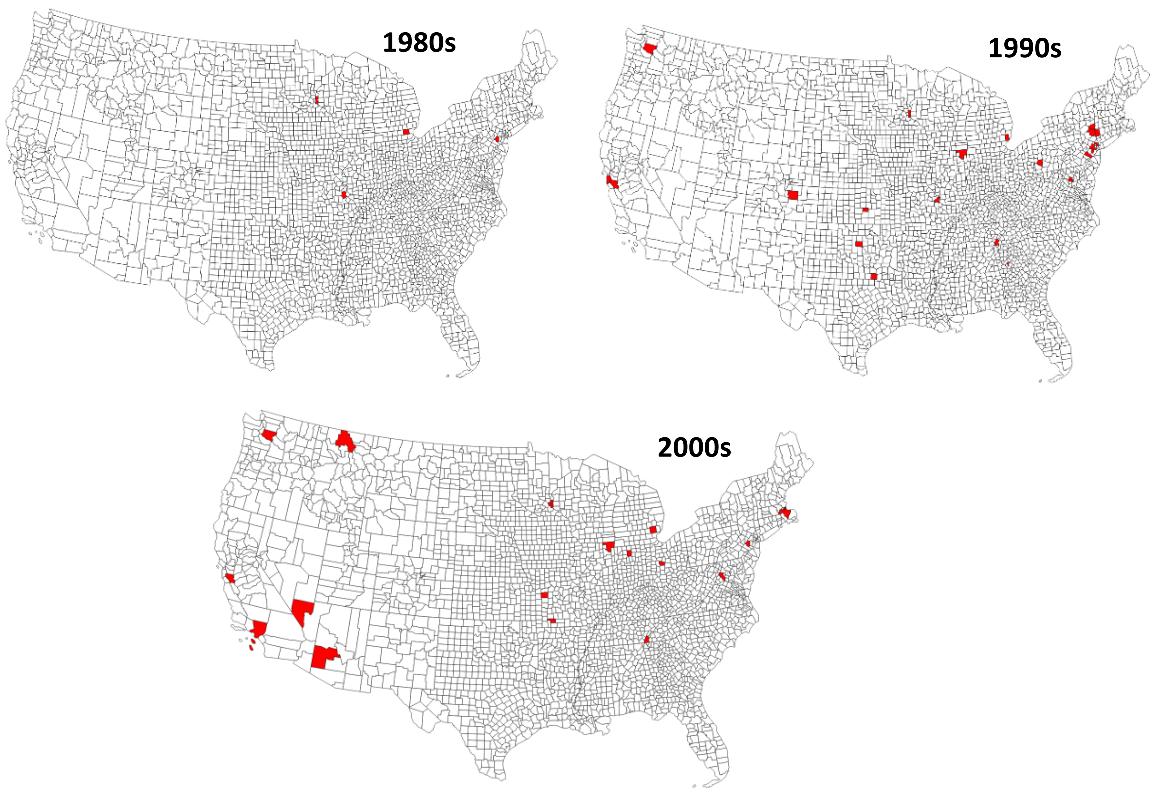

Figure 9 - Origin counties of moving inventors within the U.S. named Fleming

TABLES

Table 1 – descriptive statistics at U.S. county level, N=65,247

Variable	mean	median	std dev	min	max
Number of incoming inventors	2.15	0.00	11.96	0.00	700.00
Instrument	1.98	0.70	8.68	0.00	356.03
Number of overall venture-backed startups	0.42	0.00	4.20	0.00	314.00
Number of successful startups (RoR ≥ 125%)	0.04	0.00	0.60	0.00	38.00
Number of successful startups (RoR ≥ 500%)	0.02	0.00	0.28	0.00	19.00
Number of successful startups (Exit ≥ 1B)	0.00	0.00	0.05	0.00	4.00
Number of failed startups	0.05	0.00	0.83	0.00	91.00
Number of failed startups (inc. RoR < 125%)	0.08	0.00	0.08	0.00	123.00
Number of high-tech startups	0.34	0.00	3.79	0.00	306.00
Number of low-tech startups	0.09	0.00	0.70	0.00	38.00

Notes: This table reports summary statistics of the key variables used in our regression analyses at the county level, covering 3107 counties 1987-2007. "Successful" startups are those that complete either an IPO or successful acquisition within 10 years, and we have three different cutoffs at an exit value ≥ 125%, 500% of total venture capital acquired or an absolute exit value ≥ 1B dollars. "Failed" startups are those that are currently "Defunct" or "Bankruptcy" as indicated in VentureXpert database. In addition, we have another variable for "Failed" startups that includes startups that complete either an IPO or successful acquisition within 10 years, but achieve a value < 125% of total venture capital acquired. High- vs. low-tech startups are categorized according to VentureXpert classifications.

Table 2 – Destination county choice

	origin-destination county move			
	a	b	c	d
Destination county Historic surname fraction	0.044*** (0.006)	0.021*** (0.002)	0.044*** (0.006)	0.013*** (0.001)
N	524,583,139	524,583,139	524,583,139	523,553,217
Year FEs	Yes	Yes	Yes	Yes
Destination county FEs	No	Yes	No	No
Origin county FEs	No	No	Yes	No
Origin-destination county FEs	No	No	No	Yes
R^2	0.000	0.008	0.000	0.061

Notes: This table presents OLS regressions of a dummy indicating an origin-destination county move of an inventor within the period 1980-2015 on destination counties' historic surname shares in 1940. Unit of observation is the origin-destination county dyad. Standard errors clustered at the destination county appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 3 – Impact of incoming inventors on local venture backed startups

	Venture-backed startups founded				
	a	b	c	d	e
	OLS	IV	IV	IV	IV (w/o top 10 counties)
Incoming Inventors$_{t-1}$	0.360***	0.510***	0.513***	0.180***	0.170***
	(0.019)	(0.027)	(0.027)	(0.040)	(0.040)
N	65,247	65,247	65,247	65,247	65,058
First stage F		804.265	781.375	175.723	139.252
Year FE	Yes	Yes	No	No	No
State FE	Yes	Yes	No	No	No
State-Year FE	No	No	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes
R^2	0.500				

Notes: This table presents OLS regressions of log (number of venture-backed startups + 1). Incoming inventors as well as the instrument are log-transformed. Specifications (b)-(d) show results of our IV regression as described above, where incoming inventors are instrumented with $Inv_{dt,leave-out}$ in the first stage. Specification (e) show results of our IV regression, but excluding top 10 entrepreneurial counties from the sample. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 4 – Industry-specific inventors and startups

	Venture-backed startups founded	
	a	b
	In same industry	In different industries
	IV	IV
Incoming Inventors$_{t-1}$	0.507***	-0.320***
	(0.052)	(0.033)
N	326,235	326,235
First Stage F	143.955	143.955
County-Industry FE	Yes	Yes
County-Year FE	Yes	Yes
Industry-Year FE	Yes	Yes

Notes: This table presents OLS regressions of log(number of venture-backed startups + 1). All specifications show results of our IV regression as described above, where incoming inventors are instrumented with $Inv_{dt,leave-out}$ in the first stage. Specifications (a) and (b) present the results for number of venture-backed startups founded in the same and different industries compared to the expertise of incoming inventors, respectively. Incoming inventors are instrumented with $Inv_{dt,leave-out}$ in the first stage. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 5 – Venture-backed startups: Successful vs. Failure

	Successful venture-backed startups			**Failed venture-backed startups**	
	a	b	c	d	e
	Successful (RoR ≥ 125%)	Successful (RoR ≥ 500%)	Successful (Exit ≥ 1B)	Failed	Failed or RoR < 125%
	IV	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	0.104***	0.068***	0.014**	-0.212***	-0.123***
	(0.033)	(0.023)	(0.006)	(0.028)	(0.027)
N	65,247	65,247	65,247	65,247	65,247
First Stage F	175.723	175.723	175.723	175.723	175.723
State-Year FE	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: This table presents OLS regressions of log(number of venture-backed startup foundations + 1). All specifications show results of our IV regression as described above, where incoming inventors are instrumented with $Inv_{dt,leave-out}$ in the first stage. In specification (a), we define "successful" startups as those that complete either an IPO or successful acquisition within 10 years and achieve a value ≥ 125% of total venture capital acquired. In specification (b), we raised the threshold of an exit value to 500% of total venture capital acquired. In specification (c), we define we define "successful" startups as those that complete either an IPO or successful acquisition within 10 years and achieve an absolute value ≥ 1B dollars, respectively. In specification (d), we define "failed" startups as those that are currently "Defunct" or "Bankruptcy" as indicated in VentureXpert database. In specification (e), we also include startups that complete either an IPO or successful acquisition within 10 years, but achieve a value < 125% of total venture capital acquired. Incoming inventors as well as the instrument are log-transformed. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 6 – Venture-backed startups: high-tech vs low-tech, successful vs. unsuccessful

	Low tech			High tech		
	a	b	c	d	e	f
	All startups	Successful	Failed	All startups	Successful	Failed
	IV	IV	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	-0.136***	-0.017*	-0.163***	0.356***	0.128***	-0.083***
	(0.029)	(0.010)	(0.022)	(0.042)	(0.034)	(0.018)
N	65,247	65,247	65,247	65,247	65,247	65,247
First Stage F	175.723	175.723	175.723	175.723	175.723	175.723
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents OLS regressions of log(number of startup foundations + 1) separated by high tech and low tech industries. High- vs. low-tech are categorized according to VentureXpert classifications. Specification (a) and (d) show results of all venture-backed startups foundations. Specification (b) and (e) show results of successful venture-backed startups foundations, where "successful" startups are defined as newly founded venture-backed startups that complete either an IPO or successful acquisition within 10 years and achieve a value ≥ 125% of total venture capital acquired. Specification (c) and (f) show results of failure venture-backed startups foundations, where "failed" startups are defined as those that are currently "Defunct" or "Bankruptcy" as indicated in VentureXpert database. Incoming inventors as well as the instrument are log-transformed. All specifications show results of our IV regression as described above, where incoming inventors are instrumented with $Inv_{dt,leave-out}$ in the first stage. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 7 – Alternative instruments

	Successful venture-backed startups founded			
	a	b	c	d
	Only individuals settled by 1935	Dropped 50 most frequent surnames	Dropped wealthy families	Alternative instrument using centroid
	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	0.110***	0.109***	0.106***	0.272**
	(0.036)	(0.035)	(0.033)	(0.109)
N	65,247	65,247	65,247	65,247
First Stage F	159.068	162.303	183.076	22.971
State-Year FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes

Notes: This table presents OLS regression of log(number of successful venture-backed startups founded + 1), where "successful" startups are defined as newly found venture backed companies that complete either an IPO or successful acquisition within 10 years and achieve a value > 125% of total venture capital acquired. Incoming inventors as well as the instrument are log-transformed. Model (a) restricts the instrument to those who settled in the county of the 1940 Census by 1935; (b) excludes the 50 most frequent surnames; (c) excludes the wealthiest 1% of surnames per 1940 Census house value; (d) replaces the shift-share approach with the inverse geographic distance between the county and the centroid for the inventor's surname. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 8 – Results from placebo analysis

	a	b	c	d
	Coefficient		Std. Err.	Rejection rate
	(Mean)	(Std. Dev.)	(Median)	(%)
Panel A: Placebo IV randomly shuffled across the overall sample				
1st stage	0.000	0.002	0.002	5.5
2nd stage	2.223	67.497	0.598	0.0
Panel B: Placebo IV randomly shuffled across counties within each year				
1st stage	0.000	0.002	0.002	5.2
2nd stage	0.090	11.871	0.638	0.1
Panel C: Placebo IV randomly shuffled across years within each county				
1st stage	-0.004	0.008	0.007	8.5
2nd stage	-0.264	4.164	0.742	0.1

Notes: We randomly shuffle our instrument to construct placebo instrument variables across the overall sample (Panel A), across counties within each year (Panel B), and across years within each county (Panel C). For each placebo instrument variables, we ran 1000 regressions of ln (number of successful venture-backed startup foundation + 1) on incoming inventors, instrumented with the placebo IV that is newly generated for each regression. Incoming inventors as well as the placebo instrument are log-transformed. Column (a) and (b) report the mean and standard deviation of the coefficients obtained from 1000 placebo regressions, respectively. Column (c) reports the median value of the standard error for the coefficient of each regression over 1000 placebo regressions. Column (d) reports the rate of which the regression rejects the null hypothesis of no effect at the 5% significance level over 1000 placebo regressions. We report these values corresponding to each of the first and second stages of the placebo regressions.

APPENDICES

Appendix A1: Matching between surnames in patent and Census data

Matching surnames between Census and patent data requires cleaning of the surname raw strings. We convert all surnames to lower cases and delete unnecessary punctuations and other noise in the surnames (e.g., ' "_ / & ; ( ) - =). We also remove suffixes and other extra words after commas (e.g., 'Foster', 'Sr.', 'deceased'). This process reduces unique surname strings down to 3,313,643 unique surnames in the Census data and 330,098 unique surnames in the patent data. Out of 374,988 inventor surname raw strings, a total of 275,849 (73.6%) find a match in the census surname. Compared to the matching without these cleaning processes, which finds 230,421 census surname matches out of 374,988 inventor surname raw strings (61.4%), our name cleaning process adds 12.2% of matches. In our data sample specifically, out of 3,165,207 unique inventors that applied for at least one patent in US, 2,894,917 inventors (91.5%) match their surname to the Census data.

Appendix A2: Disambiguating geographic location and matching to a county

Although most U.S. patent front pages provide strings for the hometown and state of each inventor, much work must be done to accurately map those strings to counties. Figure A1 illustrates the geographic disambiguation process. We begin with updated data processed via Balsmeier et al. (2018) methods, from 1976 to 2018, which includes 16,215,831 "patent-inventor pairs" because many inventors have multiple patents. Exclusion of non-U.S. and entirely missing data fields leaves 8,065,290 U.S. patent-inventor data points. Amongst these there are 72,122 unique city-state pair strings. Note that this number includes misspellings, neighborhoods and unincorporated areas that do not correspond to city and state, and outright errors.

We exactly matched 27,299 city-state data points for 7,718,350 patent-inventors using the SimpleMaps (https://simplemaps.com/) concordance. We took the remaining unique and unmatched locations and ran them through the Google Geocoding API (https://developers.google.com/maps/documentation/geocoding/overview). This left 10,413 unique city-state pairs and 85,046 patent-inventor pairs, which manual inspection revealed to be mainly errors. 7,980,244 patent inventor pairs were ultimately matched to a city and state, for a 98.9% match rate.

Given that our instrumentation and analysis is at the county level, we need to next map city-state locations to counties. This is complicated by the fact that our data span 1940-2018 and that there have been minor changes to this mapping over time. To address this, we begin with U.S. census records of changes from 1970 to present: (https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2010.html). Then, we manually search for changes between 1940 and 1969. We incorporate substantial changes to counties such as county consolidation, part annexation, and FIPS code changes. We build a transitive association file which tracks the changes and anchors all historic changes to the 2020 SimpleMaps concordance (file will be posted upon publication). The 1940 Census doesn't cover VI (Virgin Islands), PR (Puerto Rico), AK (Alaska), and HI (Hawaii), hence, these locations are not considered in the analyses.

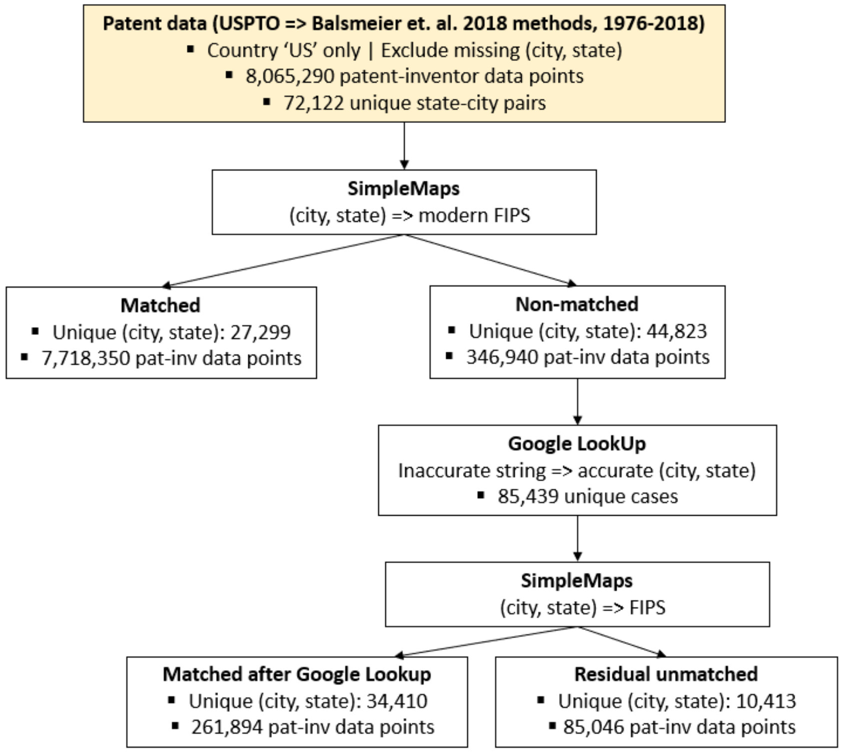

Figure A1 – Geographic disambiguation process for U.S. inventor city and state

Appendix A3: Mapping patent classes and inventors to VentureXpert categories

To estimate the impact of the influx of technology specific inventors on the startup activities of their corresponding industry, we matched NBER technological categories provided by Hall et al. (2001) with VentureXpert industry categories. Table A1 details the manual mapping of NBER technological categories to VentureXpert's major industry groups, i.e., Biotechnology, Communications and Media, Computer Related, Medical/Health/Life Science, Semiconductors/Other Elect, and Non-High-Technology. As underlying technologies overlap between the Biotechnology and Medical/Health/Life Science industry groups, we merged the two industry groups. As VentureXpert does not have corresponding industry groups for mechanical and chemical NBER technological categories, we excluded patent classes corresponding to these technological categories.

Using the concordance between VentureXpert industry groups and NBER patent classification, we classified inventors into each of the five industry groups based on the most frequent industry group that each inventor had patented in. In case of a tie, we took the earliest industry group. We excluded inventors who patented only in patent classes without a corresponding VentureXpert industry group. As a result, out of 763,715 U.S. inventors who had more than two granted patents (whose mobility could be tracked), we were able to assign 602,971 inventors to each of the five VentureXpert industry groups.

Table A1 – Concordance between VentureXpert industry groups and NBER patent classification

Industry (VentureXpert)	Sub-Category Code	Sub-Category Name	Patent Classes
Biotechnology + Medical/Health/Life Science	31	Drugs	424, 514
	32	Surgery & Medical Instruments	128, 600, 601, 602, 604, 606, 607
	33	Biotechnology	435, 800
	39	Miscellaneous-Drug & Med	351, 433, 623
Communications and Media	21	Communications	178, 333, 340, 342, 343, 358, 367, 370, 375, 379, 385, 455
Computer Related	22	Computer Hardware & Software	341, 380, 382, 395, 700, 701, 702, 704, 705, 706, 707, 708, 709, 710, 712, 713, 714
	23	Computer Peripherals	345, 347
	24	Information Storage	360, 365, 369, 711
Semiconductors/Other Elect	41	Electrical Devices	174, 200, 327, 329, 330, 331, 332, 334, 335, 336, 337, 338, 392, 439
	42	Electrical Lighting	313, 314, 315, 362, 372, 445
	43	Measuring & Testing	73, 324, 356, 374
	44	Nuclear & X-Rays	250, 376, 378
	45	Power Systems	60, 136, 290, 310, 318, 320, 322, 323, 361, 363, 388, 429
	46	Semiconductor Devices	257, 326, 438, 505
	49	Miscellaneous-Elec	191, 218, 219, 307, 346, 348, 377, 381, 386
Non-High-Technology	61	Agriculture, Husbandary, Food	43, 47, 56, 99, 111, 119, 131, 426, 449, 452, 460
	62	Amusement Devices	273, 446, 463, 472, 473
	63	Apparel & Textile	2, 12, 24, 26, 28, 36, 38, 57, 66, 68, 69, 79, 87, 112, 139, 223, 450
	64	Earth Working & Wells	37, 166, 171, 172, 175, 299, 405, 507
	65	Furniture, House Fixtures	4, 5, 30, 70, 132, 182, 211, 256, 297, 312
	66	Heating	110, 122, 126, 165, 237, 373, 431, 432
	67	Pipes & Joints	138, 277, 285, 403
	68	Receptacles	53, 206, 215, 217, 220, 224, 229, 232, 383
	69	Miscellaneous Others	1, 14, 15, 27, 33, 40, 52, 54, 59, 62, 63, 84, 101, 108, 109, 116, 134, 135, 137, 150, 160, 168, 169, 177, 181, 186, 190, 199, 231, 236, 245, 248, 249, 269, 276, 278, 279, 281, 283, 289, 292, 300, 368, 404, 412, 428, 434, 441, 462, 503