NBER WORKING PAPER SERIES

STARTUPS, UNICORNS, AND THE LOCAL INFLOW OF INVENTORS

Benjamin Balsmeier
Lee Fleming
Matt Marx
Seungryul Ryan Shin

Working Paper 27605
http://www.nber.org/papers/w27605

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020, Revised July 2024

The authors thank Guan Cheng Li for invaluable research assistance. We also thank participants in the NBER Productivity Seminar, the Munich Summer Institute 2022, the 2021 RCEA Future of Growth Conference, and in particular Lucy Xiaolu Wang 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. Shin acknowledges that this work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2024-00348787). Errors and omissions are ours. Physical address for corresponding author: Lee Fleming, Fung Institute for Engineering Leadership, 2451 Ridge Rd, Berkeley, CA 94709, 617 905 8346. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Startups, Unicorns, and the Local Inflow of Inventors
Benjamin Balsmeier, Lee Fleming, Matt Marx, and Seungryul Ryan Shin
NBER Working Paper No. 27605
July 2020, Revised July 2024
JEL No. J24,J61,L26

ABSTRACT

We provide evidence that the arrival of technical human capital improves regional entrepreneurship, both by increasing firm entry and reducing failure. The results also indicate negative externalities upon low-tech and competing industries: the arrival of inventors in a county shifts the locus of venture capital investment away from low-tech startups to high-tech startups and moreover towards new ventures in the same sector as those inventors' skills. Identification is provided by a shift-share instrument combining the spatial distribution of surnames in the 1940 U.S. Census with thousands of surname-specific shifts based on modern inventor mobility.

Benjamin Balsmeier
Department of Economics and Management
Université du Luxembourg
6, rue Richard Coudenhove-Kalergi
L-1359 Luxembourg
benjamin.balsmeier@uni.lu

Lee Fleming
330B Blum Center
Engineering Leadership
University of California, Berkeley
Berkeley, CA 94720-0001
United States
lfleming@berkeley.edu

Matt Marx
Cornell University
137 Reservoir Avenue
Ithaca, NY 14853
and NBER
mmarx@cornell.edu

Seungryul Ryan Shin
137 Reservoir Avenue
Ithaca, NY 14850
United States
s.ryan.shin@gmail.com

Entrepreneurship—especially 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). Newly founded firms are responsible for job creation (Decker et al., 2014; Glaeser, et al. 2015), productivity (Gennaioli et al, 2013) and innovation (Kortum & Lerner, 2001). Policymakers worldwide have sought to spur startup activity, 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 regarding what drives entrepreneurship. Further, given that the vast majority of new firms fail (Haltiwanger et al., 2013)—including 75% of venture-capital backed firms (Hall & Woodward, 2010)—what are the critical inputs and mechanisms that increase the rate of successful startups?

Scholars have long observed that human capital, including technical talent, is an important ingredient in the entrepreneurial recipe, but causal evidence and mechanisms of how technical human capital improves entrepreneurship remains scarce. 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).[1] Larger-scale, if suggestive, evidence for the role of inventors in successful 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

[1] 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; Guzman and Stern, 2020).

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). In this paper, we seek causal evidence on this point, in the context of high-growth startups that raise venture capital.

We begin by establishing an association between the arrival of inventors in a focal county and a subsequent increase in venture-backed startups in that county. Moreover, we find that those incoming inventors play direct roles in more than 1 of 13 new, VC-backed startups: as a founder, executive, or an inventor on the startup's patent. Even though we find linkages between inbound inventors and new startups, these correlations do not address reversal causality. That is, does the arrival of inventors in a region drive entrepreneurial activity, or are inventors more attracted to regions where they expect higher levels of entrepreneurship? Our approach to providing a more causal assessment relies upon a shift-share instrumental variable (SSIV). The "share" is the proportion of people in 1940 with a particular last name residing in each of 3,097 counties. The "shift" is the number of inventors with a particular last name who move anywhere in the U.S. each year from 1987-2007. For each county-year observation, the instrument is computed by summing over all surnames the 1940 share for that surname in the focal county multiplied by the shift for that surname in the focal year (less the number of inventors of that surname who moved to the focal county).

We employ our SSIV to estimate the impact of an inflow of inventors to a focal county on the quantity and quality of venture-capital backed startups in the county. We focus on venture financed startups; although only 0.5% of new businesses obtain venture financing (Puri & Zarutskie, 2009), nearly half of firms that complete an IPO had raised venture capital (Lerner & Nanda, 2020). We combine data from VentureXpert, Crunchbase, and Pitchbook on startups that received venture capital funding as early as 1987 and exited as late as 2020.

Our results suggest that inbound inventors improve both the quantity and quality of entrepreneurship. More venture-backed startups are founded in counties where more inventors arrive (the arrival of approximately 28 inventors implies one additional startup). Second, inbound inventors yield more startups that eventually yield an attractive return on investment for the investor (based on acquisition/IPO values), including billion-dollar “unicorn” exits. Third, inbound inventors reduce the number of startups that go bankrupt or are sold in a “fire sale.” Therefore, the supply of inventors appears to increase not only the rate but also the efficiency of entrepreneurship.

Our results suggest two ways in which inventors improve the efficiency of entrepreneurship. First, the arrival of technical talent shifts venture dollars away from low-tech entrepreneurship to high-tech entrepreneurship. Moreover, investments in low-tech startups that fail are displaced by investments in high-tech startups that succeed. Second, this displacement happens not only at the level of technical vs. non-technical startups but also at the level of sector-specific skills as the arrival of inventors in one sector depresses startup founding rates in other sectors.

We interpret these results cautiously, given the nature of our analysis. Although we attempt to validate both the shift and share components of the instrument, the SSIV remains vulnerable to at least two issues. First, because we estimate a local average treatment effect, it may be that our overall results are driven by counties with more entrepreneurial activity as well as more productive and mobile inventors. We still observe that inbound inventors result in more startups when removing Silicon Valley and other startup hotspots, albeit in smaller magnitude. Thus, we caution against inferring that simply moving inventors to a rural county with little innovative or entrepreneurial activity will suddenly yield a venture-backed startup. Second, certain types of simultaneous, interstate shocks to industry demand may yield spatial correlations that threaten inference. For this reason, we see our work as a step toward causal evidence.

The remainder of the paper is organized as follows. Section 1 describes the data. Section 2 defines the shift-share instrument. Section 3 presents basic results. Section 4 explores mechanisms, including sector-specific results. Section 5 concludes. Online appendices detail data construction and instrument plausibility and validation.

1. DATA

We assemble three different sources at varying degrees of aggregation and times to arrive at a panel dataset at the U.S. county-year level for 1987-2007.

1.1 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 explained below, our identification strategy relies on being able to observe the name and location of each U.S. citizen in 1940 in order to predict the mobility of modern inventors. The historic data include 3,363,932 different surnames, for which the median number of occurrences is 3, the mean is 39, and the maximum is 1,359,079 (for Smith). 27% appear only once. The 1940 U.S. Census records consist of 3097 counties and other districts based on the county system in 1940. In order to match the location information of inventors, we translate 19 counties or districts that are 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 A1 and A2 for details on name matching and geographic disambiguation, respectively.

1.2 Inventor data

We begin with raw data from the United States Patent and Trademark Office (USPTO) from 1976-2018 (only the overlapping time periods 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 name cleaning and standardizing procedures and the matching algorithm, we match 91.1% of inventors' surnames to a surname from the 1940 Census.

We use the inventor identification number (ID) and location provided in Balsmeier et al. (2018) to identify inventor moves across U.S. counties. (We cannot observe an inventor move with fewer than two patents.) Using the 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 intermediate 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.

1.3 Entrepreneurship data

Our main data source of US entrepreneurial activity is VentureXpert, which is part of Thompson's economic data suite and covers all venture-backed firms in the 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. Because VentureXpert is sometimes missing capital investment, acquisition values, and founding year (and some IPO values), we fill these in using Pitchbook and Crunchbase, via exact match on website URL and state (Dorn et al. 2020).

Our baseline sample consists of all startups with information on founding year, industry and location, starting in 1987 (VentureExpert lacks comprehensive coverage beforehand). We use the year of founding in our estimations, rather than the year of funding, and 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. We focus on venture-backed startups as they are important drivers of economic dynamism, innovation, and long-term growth (Decker et. al. 2014, Lerner & Nanda, 2020).

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 consider a 500% Rate of Return (RoR) on invested capital as well as "unicorns" i.e., startups which exit with a valuation of $1B or greater (independent of capital raised). For failed startups, rely on VentureXpert's classification as

"Defunct" or "Bankruptcy." Where a startup had not exited within ten years of founding, it was not counted as having failed or succeeded. Table 1 provides descriptive statistics 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. VentureXpert lists 3386 venture-backed startups as "Defunct" or "Bankruptcy."

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 at the county-year level, covering 3107 counties 1987-2007. "Successful" startups are those that complete an IPO or successful acquisition within 10 years, with three different cutoffs at an exit value ≥ 125%, 500% of total venture capital acquired or an absolute exit value ≥ 1B dollars. "Failed" startups are "Defunct" or "Bankruptcy" in VentureXpert. Another variable for "Failed" startups that includes startups with exit value < 125% of invested capital. High- vs. low-tech startups are categorized according to VentureXpert (Appendix A3).

2. CONSTRUCTING THE SHIFT-SHARE INSTRUMENT

Estimating the causal impact of mobile inventors on local startup formation is challenging due to potentially unobservable confounders at the individual level (e.g., personal reasons for moving to a certain region) as well as at the regional level (e.g., local job market prospects). We approach this with a shift-share instrumental variable (SSIV) based on Bartik (1991) and its applications to international migration (Card (2009); Burchardi et al. (2020)). Our "shift" is the total number of nationwide surnames of mobile inventors in a given year. Our "share" uses the proportion of the same surnames in a county in the 1940 U.S. Census.[2] We discuss first stage, share, and shift validity (Goldsmith-Pinkham et al. 2020; Borusyak et al. 2022) in Appendix C and D.

Prior work finds that immigrants locate near previous immigrants from the same country of origin (Bartel, 1989; Lalonde & Topel, 1991). Card (2001) and others (Jaeger, Ruist & Stuhler, 2018 has an overview) exploited this observation to predict immigrant inflows into particular U.S. regions, interacting past shares of immigrants from an origin country to a given region with the contemporaneous 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 particular county based on the spatial distribution of U.S. surnames across counties in 1940 and the total number of mobile inventor surnames in a given modern year. The intuition is thus: 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 for the full population of U.S.

[2] Guided by recent literature, we address concerns related to both the share (Goldsmith-Pinkham et al. 2020) and shift (Borusyak et al. 2022). While both approaches highlight that causality is sufficiently established if either the share or the shift is conditionally exogenous, we check that our approach is remain plausible given either emphasis.

inventors, we borrow the 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 et. al. 2011, for the example of Savoy, France and Clark & Cummins 2015, for England). This literature also found that historic locations of family members can serve to predict movements of people decades later (Darlu et. al. 2011; Clark & Cummins 2015). Appendix C verifies that these patterns hold for individual U.S. inventors.

Our approach departs from Card (2009) in two ways. First, resembling recent advances in international migration literature (Buchardi et al. 2020, Hunt 2017, and Wozniak & Murray 2012) we try to avoid unobserved local demand shocks that might invalidate the instrument by removing each county's own inventor inflows from the nationwide shifts (see below regarding the leave-out extension of the instrument). Second, by utilizing more than 200,000 surname-level shares, our approach minimizes reliance upon the influence of any particular share as can happen with nation-level shares. Ideally, the historic share of any particular name in a certain county should not be correlated with unobserved factors that explain why counties with larger shares of that name in 1940 will generate larger *changes in future entrepreneurship*.

The intuition underlying 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. Although we employ a larger number of shares than most prior SSIV applications, the instrument remains vulnerable to criticisms including that a local demand shock in a particular country could drive inbound inventors and entrepreneurship. Our preferred SSIV specification addresses one case of local shocks via a "leave-out" approach, and we discuss other cases (and remaining limitations) below and in Appendix D.

2.1 Definition of the SSIV

In the absence of endogeneity concerns, we could formally estimate the effect of inventor mobility on entrepreneurship via OLS:

$$Y_{ds,t} = \alpha_0 + \beta \cdot Inv_{ds,t-1} + \delta_t \times \eta_s + \gamma_d + \varepsilon_{dst} \qquad (1)$$

where $Y_{ds,t}$ is a dependent variable observed for county *d* in state *s* in year *t*. $Inv_{ds,t-1}$ is the number of inventors who moved to county *d* in state *s* in the previous period *t-1*. As explained in the data section above, an inward inventor move is observed from patent documents, leading to a fuzzy lag of at least one year (this assumption is robust to estimations of other lags). δ_t denotes year fixed effects and η_s denotes state fixed effects. We control for unobserved time variant and time invariant state specific shocks, such as varying state-level economic conditions and policy changes, through state-year fixed effects $\delta_t \times \eta_s$. γ_d controls for time-invariant unobserved county characteristics that may confound our identification of β. ε_{sdt} 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 control for any persistent differences in innovation levels across counties, this misses temporary local shocks that might attract inventors. The SSIV approach aims to address this econometric challenge by constructing an instrument that is orthogonal to (or uncorrelated with) time-varying unobservable confounders.

We define our instrument as:

$$\widetilde{Inv_{d,t-1}} = \sum_n \frac{P_{dn}^{1940}}{P_n^{1940}} \cdot Inv_{n,t-1} \qquad (2)$$

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 $Inv_{n,t-1}$ 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-1*. The expected inflow of inventors $\widetilde{Inv_{dt-1}}$ in county *d* at time *t-1* is thus the weighted sum of inventors that move across the U.S. with surname *n* at time *t-1* (the "shift") with the historical distribution of the same family names (the "shares") serving as weights.

Another concern might 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 Marx 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 Marx 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, and Borusyak et. al. 2022 for theoretical explanation and proofs). Our preferred instrument is thus:

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

where $Inv_{n,t-1,leave-out(n,d)}$ is the total number of inventors with name *n* who move to counties outside of *d*. The leave-out strategy aims to avoid the possibility that potentially demand driven choices, e.g., of Marxes, to move to county *d*, do not drive changes in the instrument.

2.2 Validation and caveats

Appendix D investigates additional threats to SSIV identification, including correlation with modern pull factors (D1), the possibility that certain surnames are drawn to economically advantaged counties (D2), and the influence of popular or wealthy surnames (D14). Two points deserve particular attention. First, because the instrument estimates a local average treatment effect across all counties, it could be that counties with higher levels of entrepreneurship are overweighted, as are productive inventors, whose mobility can be more precisely measured because they have more patents. This could lead to larger marginal effects than we should expect for the average county or inventor. Appendix D14 removes the top 5% of counties according to the SSIV, with directionally consistent results at similar significance levels yet smaller magnitudes. Therefore, our estimates should not be viewed as applying to every county, especially those with little prior activity in innovation or entrepreneurship.

The second point returns to the issue of demand shocks. Although the leave-out modification to the instrument helps to address an isolated demand shock, certain types of simultaneous, industry-specific shocks in multiple counties that are spatially separated might nonetheless present a challenge. The industry-level version of the instrument we calculated in Section 4.2 helps to address this, but in Appendix D8 we describe in detail a scenario (requiring several jointly-held conditions) that may yield unresolved spatial correlation.

3. RESULTS

Table 2 provides baseline results for the impact of inbound inventors on entrepreneurship. These 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*. We begin with correlations and then move to the instrument defined in Section 2.

Model (a) of Table 2 estimates Equation (1), including year fixed effects to help absorb unobserved macroeconomic changes. This cross-sectional analysis reveals a strong correlation between the number of incoming inventors in a county with the count of venture-backed startups founded in the next period, consistent with Kerr (2013), Maloney & Caicedo (2016), and Azoulay et al. (2020). Model (b) includes state-year fixed effects, suggesting little impact from unobserved, state-based policy changes. Model (c) adds county fixed effects. The substantially smaller estimated magnitude in model (c) indicates that unobserved time-invariant confounders explain a large part of the raw correlation between incoming inventors and entrepreneurship.

Table 2 – Impact of incoming inventors on local venture backed startups

	# VC-backed startups			Incoming Inventors$_{t-1}$	# VC-backed startups	
	a	b	c	d	e	f
	OLS	OLS	OLS	OLS (first stage)	IV	IV (w/o top 10 counties)
$\widetilde{Inv_{d,t-1,leave-out}}$				0.309***		
				(0.023)		
Incoming Inventors$_{t-1}$	0.358***	0.363***	0.031***		0.180***	0.119***
	(0.019)	(0.019)	(0.005)		(0.040)	(0.036)
N	65,247	65,247	65,247	65,247	65,247	65,058
First stage F					175.723	170.535
Year FE	Yes	No	No	No	No	No
State-Year FE	No	Yes	Yes	Yes	Yes	Yes
County FE	No	No	Yes	Yes	Yes	Yes
R^2	0.482	0.509	0.835	0.861		

Notes: Models (a-c) present OLS regressions of log (# venture-backed startups + 1). Incoming inventors and the instrument are log-transformed. Model (a) includes year fixed effects; (b) includes state-year fixed effects; (c) adds county fixed effects. Model (d) presents first-stage results of incoming inventors on the IV (i.e. $\widetilde{Inv_{d,t-1,leave-out}}$). Model (e) shows our IV regression with state-year and county fixed effects, where incoming inventors are instrumented with $\widetilde{Inv_{d,t-1,leave-out}}$ in the first stage as in equation (3). Model (f) shows our IV without the top 10 entrepreneurial counties (Alameda, Los Angeles, Orange, San Diego, San Francisco, San Mateo, Santa Clara in California; Middlesex in Massachusetts; New York County in New York; King in Washington). First stage F is the Kleibergen-Paap Wald F statistic. Standard errors clustered at the county level are in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

The remaining columns of Table 2 explore the instrument. Model (d) reports a strong first stage, which is fully characterized in Appendix C where we estimate a dyad model measuring each inventor's mobility to the focal county as well as all other possible counties, net of the surname shares used to calculate the instrument. Models (e-f) estimate our IV approach with the instrument defined in Equation (3), including fixed effects at the state-year and county level. The first stage *F* value is in both cases well above conventional levels, suggesting that the IV regression does not suffer from weak instrument bias (Stock & Yogo, 2002; Lee et al., 2021). The instrumented regressions both indicate a significant positive impact of incoming inventors in a given county on the local rate of startup formation. Importantly, the effect is not limited to entrepreneurial "hotspots" such as Silicon Valley: removing the top ten countries by startup activity preserves the result (albeit with somewhat reduced economic significance, to which we return below).

Under the assumption that the estimated coefficient can be interpreted as an elasticity, Model (e) indicates that a 10% increase in the rate of incoming inventors increases the rate of venture-backed startups founded by 1.8% at the mean.[3] Translating the relative increases into absolute numbers at the mean of 2.15 incoming inventors (from Table 1) suggests that a county can expect one additional venture-backed startup for every 28.3 incoming inventors. However, these back of the envelope estimates are at the mean of the data, so they may not pertain to a particular region. For example, one should not infer from these estimates that a county that has never had a venture

[3] Given the estimation of the ln of number of startups and coefficient of 0.18 (model e), a 100% increase in the number of incoming inventors implies an 18% increase in the number of startups. At the means, this implies that 2.15 incoming inventors results in 0.423*0.18 = 0.07614 more startups. To observe one whole startup therefore implies 2.15*(1/.07614) = 28.3 additional incoming inventors. Again starting at the means for the entire dataset, the number outside of the top 10 entrepreneurial regions (model f) is 2.15*(1/(0.423*0.119)) = 42.7 additional inventors.

backed start-up can import 29 inventors and expect a VC-backed startup to appear. Interpreting the elasticity from Model (f), counties outside of the top ten would require 42.7 incoming inventors. Our LATE estimates likely overweight top counties, as well as prolific inventors—whose mobility is easier to detect from the patent record, and who might move for reasons negatively or unrelated to local entrepreneurship, e.g., tax benefits (Moretti & Wilson, 2014) or weather, leading to negatively biased OLS estimates. Another reason the OLS estimates in Model (c) are smaller than the IV in (e) could be attenuation bias stemming from a measurement error in the endogenous variable, possibly due to imprecise time stamps on inventor moves or errors in the locations mentioned on patent documents. Appendix D dives more deeply into the validation of the SSIV, with D16 exploring the difference in magnitude between OLS and IV.

3.1 Quality of startups founded

Table 2 establishes that the arrival of inventors correlates with an increase in the quantity of new firms but does not speak to their quality. Although many governments adopt the number of startups as an easy-to-count metric of entrepreneurship, startups both create and destroy jobs, because failure is the modal outcome (Lerner, 2009), Haltiwanger et. al. (2013). More nuanced measures of quality would be preferable, though how to measure "success" is not obvious. Although Initial Public Offerings almost always indicate a successful startup, acquisitions can be an ambiguous indicator of success. For example, Puri & Zarutski (2012) report that many venture-backed failures are "disguised" as acquisitions, often sold for pennies on the dollar. Hence, we develop more nuanced success measures using financial data on exits.

In Table 3 we first consider the venture-backed startups founded in county *d* during year *t* that become successful within a ten-year window as the dependent variable.[4] 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% (as in Table 2, at the mean).

The result in Model (a) indicates that incoming inventors are not only responsible for an increase in entrepreneurial activity, as in Table 2, but also in successful startups. One might wonder whether these inventors are only responsible for startups that "just barely" succeeded in returning capital to investors, as opposed to generating more spectacular returns. We raised the threshold of an exit value to 500% of invested capital in model (b), which reduces the magnitude of the estimated coefficient but remains statistically significant. In model (c), we show that the arrival of inventors even appears to increase "unicorn" startups with exit values in excess of 1 billion dollars, albeit with much smaller economic magnitude.

Of course, this increase in the number of successful startups could be a mechanical result of "more shots on goal." That is, investors who place more bets on more startups could succeed more often, even if the odds of success remain unchanged. Therefore, we also test how the inflow 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 a traditional measure of "failed" startups as those that are currently Defunct or Bankrupt as indicated in VentureXpert. Model (d) suggests that incoming inventors reduce the formation of failed startups in the county.

---

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

Mindful of Puri & Zarutskie (2012)'s observation that many failed venture-backed startups are "disguised" as acquisitions, in Model (e) we include as failures any exit with a valuation lower than 125% of total of invested capital. Model (e) likewise shows a negative effect of incoming inventors on failed startup foundings (and is robust to eliminating exits with >100% return on investment, or >50%). The results suggest that inventors not only improve the quantity but also the quality of entrepreneurship.

Table 3 – Venture-backed startups: Successful vs. Failure

	Successful VC-backed startups			Failed VC-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 ln (# venture-backed startup foundings + 1). All specifications show results of our IV regression as described above, where incoming inventors are instrumented with $\widetilde{Inv_{d,t-1,leave-out}}$ 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 and 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.

4. MECHANISMS

Our results indicate that incoming inventors increase both the rate and quality of startup formation. How and why does this happen? In this section, we explore three possible mechanisms relating to human capital. First, we observe a reallocation from venture capital investments from low-tech startups to high-tech following the arrival of inventors in a county. Second, we note further reallocation of VC investments into the sectors that match incoming inventors' skills. Third, we find that inbound inventors frequently serve as founders, executives, and technical staff for newly-founded startups in that same county.

4.1 Incoming inventors and VC reallocation from low-tech into high-tech sectors

Multiple studies suggest that venture capitalists' investment activity tends to be localized (Sorenson & Stuart, 2001; Bernstein et. al. 2016), given their role in mentoring and monitoring portfolio companies. To the extent that VCs focus primarily on local opportunities and allocate funds accordingly, in the absence of key human capital they might fund startups that do not rely on that type of expertise. We first explore this potential reallocation at a broad level, distinguishing between high-tech and low-tech startups. As described in Section 1.3 and Table 1, we define as "low-tech" those ventures not listed by VentureXpert as biotechnology, life science, computers, or communication and semiconductors.

Table 4 separates our dependent variable into investments in low-tech startups (Models a-c) vs. high-tech (Models d-f). Models (a) and (d) resemble Table 2 in using the count of startups as the dependent variable. The estimated coefficient on incoming inventors for model (a) is negative, compared with positive in model (d). This suggests a reallocation of venture investments from low-tech to high-tech startups following the arrival of inventors in a focal county, consistent with the increased availability of technical human capital.

The remaining models of Table 4 resemble Table 3 in exploring the quality of startups that received venture investments, segmented into low-tech (models b-c) vs. high-tech (models e-f). Starting with model (c) of Table 4, the negative, precisely estimated coefficient indicates a shift away from low-tech startups that fail. The evidence in Model (b) may suggest that successful low-tech startups also decrease in response to the arrival of inventors, though the estimated coefficient is much smaller in magnitude than that for failed low-tech startups and 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 are able to avoid less promising low-tech startups when more high-tech human capital is available. Models (e) and (f) largely echo the results of Table 3.

Table 4 – 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 ln (# startup foundings + 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 foundings. Specification (b) and (e) show results of successful venture-backed startups foundings, 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 foundings, 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 $Inv_{d,\widetilde{t-1,leave-out}}$ 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.

4.2 Incoming inventors precede the reallocation of investment towards their sector

Table 5 builds on the human-capital analysis of Table 4, segmenting not just between high-tech and low-tech but examining allocation of capital to sectors that match the expertise of incoming inventors. Evidence on this point would not only point to a potential mechanism but also suggest, similar to the analysis of Table 4, that increased investment in one sector comes at the expense of another. If this were not the case, one might question the baseline results. If for example a focal county only had an influx of software inventors, yet all of the increase in startup activity was in biotechnology, we might wonder whether our informal model accurately captures an application of task-specific human capital (Gibbons & Waldman, 2004) to relevant startups. This result might also decrease potential measurement error, i.e., a downward bias in the aggregated estimations of Table 2, if the arrival of inventors has a negative effect outside of their field.

To address these theoretical and econometric concerns, we created an additional dataset at the destination county-industry-year level. We differentiated between each of the four high-tech classifications and the low-tech sector as defined by VentureXpert and matched inventors with these industries based on the technology classification assigned to each patent. If an inventor patents in multiple technology classifications, we used the most frequent one. In case of a tie, we used the earliest classification (see Appendix A3 for details). We used these industry-specific classifications to create a county-industry-year version of the county-year SSIV described in Section 2. The share is unchanged (as the 1940 Census does not have industry classifications), but the shift is calculated according to the yearly number of mobile inventors with the same surname in each industry. We then estimate the following equation with OLS:

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

where $Y_{d,i,t}$ stands for a dependent variable observed for county d, industry i at time t. $Inv_{d,i,t-1}$ 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 from equation (1) is that the county-industry-year data allows us to control for unobserved time-varying, county-level characteristics via county-year fixed effects $\gamma_d \times \delta_t$. This includes for instance the total number of inventors in a county as it grows or shrinks over time or the stock of unobserved factors related to entrepreneurship. Put differently, identification of β will only come from relative differences across industries within a county and year, so we only expect β to be positive if (for example) a higher *fraction* of biotech inventors out of all inventors moving into a given county at a given time leads to a higher *fraction* of biotech startups within the same county and at the same time.

To absorb unobserved industry-specific trends, we add industry-year fixed effects $\theta_i \times \delta_t$, and to address unobserved industry-specific regional advantages we add industry-county fixed effects $\theta_i \times \gamma_d$. Since all fixed effects enter both the first and second stage of our IV regressions, this helps to alleviate concerns with respect to unobserved trends in the attractiveness of certain regions that may influence the mobility of inventors. As just one example, these fixed effects help to account for Silicon Valley's growth in demand for semiconductor engineers. The match between industry-specific human capital and industry-specific entrepreneurship should also reduce measurement error, so we expect β to be larger when estimated with county-industry-year data in equation (4) than with county-year data in (3). The finer unit of measurement leads to a larger number of observations, but the underlying data remain unchanged.

Table 5 illustrates the results of the county-industry-year instrument, estimating Equation (4). The analysis resembles that of Table 2 model (e), but the dependent variable is industry-specific. In model (a), we measure the number of startups founded in focal county *d*, industry *i*, and time *t*.

The coefficient on incoming inventors in model (a) is positive and precisely estimated, indicating that the number of startups in a given industry increases following the arrival of inventors with skills in the same industry. Table 5 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. That this estimated coefficient is larger than that of model (d) in Table 2 is consistent with the aforementioned reduction in measurement error.

Table 5 – 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 $Inv_{d,\widetilde{t-1,leave}-out}$ 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.

The field-specific nature of this exposure is further reinforced by model (b), which instead uses a dependent variable of the number VC-backed startups *outside* the focal industry (i.e., $Y_{d,-i,t}$). A positive estimate for the coefficient on Incoming Inventors in model (b) would suggest that our effect is not driven by task-specific human capital but more generic technical skills; the sign, however, is negative. This offsetting result makes sense in the context of venture-backed startups, as venture investors must decide how to allocate a fixed number of 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.

The results from Table 5 help to characterize the threat to identification raised in Section 2.2. That we can establish industry-specific results should help to allay concerns that the county-year instrument defined in Section 2 and used to estimate Tables 2-4 is vulnerable to general demand shocks. However, and as we discuss further in Appendix D8, under certain conditions simultaneous industry-specific shocks across states could still result in spatial correlation.

4.4 Incoming inventors provide founders and employees for local venture-backed startups

After establishing a link between industry-specific skills and startups in that sector, we now investigate direct linkages between inbound inventors and new startups by tracing the roles taken by newly-arriving inventors in those ventures. Perhaps most intuitively, inventors may serve as the founders of new ventures (Gambardella, Ganco, & Honore, 2015), but workers have heterogeneous preferences regarding founding a startup vs joining one (Roach and Sauermann, 2015), so we might also expect to incoming inventors among non-founding executives and engineers who yet provide crucial human capital to the venture.

We investigated direct linkages between incoming inventors and newly-founded startups by mapping the names of those inventors to those of startup personnel. The VentureXpert database had limited data regarding personnel, so we complemented these with PitchBook and Crunchbase. We found names of 169,982 founders and other non-founding executives, augmented with inventors who held patents at those ventures (by cross walking the names of patent assignees to PitchBook, 89,863 inventors total.) These combined sources yielded data on personnel for 78% of startups in our sample. We believe this to be a lower bound, given the difficulty of finding personnel for older startups as well as the fact that VentureXpert, Crunchbase, and Pitchbook generally limit their lists to founders and executives.

The next step was to compare these names of startup personnel to the names of incoming inventors. We dropped matches where the incoming inventor's first and last name were in the top 1% of patenting inventors. The crosswalk was originally built with the same lag structure of the paper but then enlarged to 3-year windows before and after the focal year, given the potential for incoming inventors to contribute before and after the company was founded. This resulted in an incoming inventor among the personnel for 7.4% of startups. (This figure grows to 8.4% if we do not exclude common-name matches and drops to 1.6% if we use only a one-year window.) Looking at founders only, we located founders for 3.8% of startups. This meaningful fraction of startups that are staffed by those who had very recently moved to the same county shows a direct link between incoming inventors and newly-founded startups.

In Table 6, we re-estimated our SSIV model with a dependent variable counting the VC-backed startups with a direct link to incoming inventors. The dependent variable is limited to startups with (a) an incoming inventor as a founder; (b) an incoming inventor as a non-founding executive; (c) as an inventor at the startup, (d) any of the above. In all models, we estimate a positive and statistically significant coefficient on Incoming Inventors. The estimated effect of all of these groups together is slightly larger than the full sample (0.183 vs. 0.180) while statistical significance remains similar, despite the smaller sample size. In sum, we find consistent results when restricting our analysis to newly-founded startups where an incoming inventor can be found among the employees of the startup.

Table 6 – Venture-backed startups with founders or employees from incoming inventors

	Startups with an executive or an inventor traced to an incoming inventor			
	a	b	c	d
	Incoming inventor was a founder	Incoming inventor was a non-founding executive	Incoming inventor held a patent at the startup	Founder, non-founding executive, or held a patent
	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	0.079***	0.121***	0.067***	0.183***
	(0.015)	(0.024)	(0.017)	(0.031)
N	65,247	65,247	65,247	65,247
First Stage F	175.723	175.723	175.723	175.723
State-Year FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes

Notes: This table presents OLS regressions of ln (number of startup foundings + 1) for restricted sample of startups with an executive or an inventor traced to an incoming inventor. Specification (a) shows the results for all venture-backed startups with a founder inventor traced to an incoming inventor. Specification (b) shows the results for all venture-backed startups with a non-founder executive inventor traced to an incoming inventor. Specification (c) shows the results for all venture-backed startups with an inventor who patented at a local startup and can be traced to an incoming inventor. Specification (d) shows the results of all venture-backed startups with an executive or an inventor traced to an incoming inventor. 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 $Inv_{\widetilde{d,t-1,leave-out}}$ 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.

5. CONCLUSION

Although a correlation between technical human capital and entrepreneurial success has often been observed (Kerr, 2013; Maloney & Caicedo, 2016; Azoulay et al., 2020), causality and mechanisms have remained elusive. Do scientists and engineers enable successful startups in their field, or do they simply flock to opportunity? We take steps toward causal inference with a shift-share instrument based on surnames from the 1940 U.S. Census and modern inventor mobility across U.S. counties, building on prior methods in the international-migration literature

(Card 2009). Our results suggest that the arrival of inventors in a county increases both the quantity and quality of local entrepreneurship by providing task-specific human capital for founders, executives, engineers, and scientists for newly founded startups. Venture investments shift towards high-tech opportunities, at the expense of unsuccessful low technology opportunities. The effect is industry-specific, with incoming inventors with industry-specific skills yielding an uptick in related startups and a corresponding decrease in unrelated industries. The results hold when Silicon Valley and similar hotspots are dropped from consideration—and when examining only inventors who are employed at startups in the focal county.

We caution that the instrument risks being influenced by multiple, simultaneous demand shocks across state lines where a significant number of common names appear. There are also limits to generalizability as we estimate local average treatment effects whereas many counties have a limited or no history of high-tech entrepreneurship. One should not infer that adding dozens (or hundreds) of inventors to a rural county would yield a unicorn in the short run. Our work should therefore be viewed as an initial step toward causal evidence of how mobility influences regional outcomes.

Although this work sought to explain how the supply of inventors influenced successful entrepreneurship, it can speak to the classic question of why industries cluster geographically (Rosenthal & Strange, 2004; Overman & Puga, 2010; Ellison et al. 2010). Much work supports the Marshallian agglomeration arguments of production economies, labor pooling, and knowledge spillovers, yet this research has struggled to isolate and estimate causal mechanisms (Glaeser & Kerr 2009; Myers and Lanahan, 2022; Balsmeier et. al. 2023). These results specifically illustrated how inventor arrival could fuel an increase and funding in startups in those inventors' specific industries – at the expense of competing high tech industries in the region. And while this paper counts startup events, it is likely that incoming inventors also help young firms scale.

While our focus has been the impact on the county in receipt of inventors, a complementary question is how the loss of inventors and other technical talent impacts the source region. Evidence on this point would inform discussions of regional inequality. Although beyond the scope of this work, a fuller accounting of these possibly-countervailing effects could enable an estimate of the social welfare of inventor mobility. Should policies encourage industries and technologies to cluster, or should industries be encouraged to disperse, and jobs distributed in a more geographically equitable way?

REFERENCES

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.

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G. Lueck, S., O'Reagan, D., Yeh, B. Zang, G., 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.

Balsmeier, B. and L. Fleming, S. Lueck (2023). "Isolating Personal Knowledge Spillovers: Co-inventor Deaths and Spatial Citation Differentials." *American Economic Review: Insights,* 5 (1): 21–34.

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?

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

Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-experimental shift-share research designs. *Review of Economic Studies:* Volume 89, Issue 1, pp. 181–213.

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.

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

Card, D. (2009). Immigration and Inequality. *American Economic Review: Papers & Proceedings,* 99:2, 1–21.

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.

de Chaisemartin, C. and X. D'Haultfoeuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." American Economic Review 110:9 pp. 2964-96.

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.

Gambardella, A., Ganco, M., & Honoré, F. (2015). Using what you know: Patented knowledge in incumbent firms and employee entrepreneurship. Organization Science, 26(2), 456-474.

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.
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.
Goodman-Bacon, A. (2021). "Difference-in-differences with variation in treatment timing." Journal of Econometrics, 2021, vol. 225, issue 2, 254-277.
Guzman, J. and S. Stern, (2020). "The State of American Entrepreneurship: New Estimates of the Quality and Quantity of Entrepreneurship for 32 US States, 1988-2014," American Economic Journal: Economic Policy 12(4): 212-243.
Hall, R. E., & Woodward, S. E. (2010). The burden of the non-diversifiable 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.
Hunt, J. (2017). The impact of immigration on the educational attainment of natives. *Journal of Human Resources*, 52(4), 1060-1118.
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.
Kenney, M., & Patton, D. (2011). Does inventor ownership encourage university research-derived entrepreneurship? A six university comparison. Research Policy, 40(8), 1100-1112.
Kerr, W. R. (2013). *US high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence* (No. w19377). National Bureau of Economic Research.
Klepper, S., & Thompson, P. (2010). Disagreements and intra-industry spinoffs. International journal of industrial organization, 28(5), 526-538.
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., & Wilson, D. J. (2014). State incentives for innovation, star scientists and jobs: Evidence from biotech. *Journal of Urban Economics*, 79, 20-38.
Myers, Kyle R., and Lauren Lanahan. 2022. "Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy." *American Economic Review*, 112 (7): 2393–2423.
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.
Roach, M., & Sauermann, H. (2015). Founder or joiner? The role of preferences and context in shaping different entrepreneurial interests. Management Science, 61(9), 2160-2184.
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.
Schumpeter, J. (1942) *Capitalism, socialism and democracy*. Vol. 36, Harper & Row, New York.
Smith, A. (1776). *Wealth of nations.* Oxford, England: Bibliomania.com Ltd, 2002.
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.
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.

Appendix A: Data construction

A1: Matching algorithm between surnames in patent and Census data

Matching surnames between Census and patent data requires cleaning of the raw surname 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%) ultimately match their surname to the Census data.

A2: Disambiguating geographic location and matching to a county

Although most U.S. patent front page data provide strings for the hometown and state of each inventor, much work must be done to accurately map those strings to counties. Figure A2 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 with no correspondence to city or state, and 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 county 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 dropped.

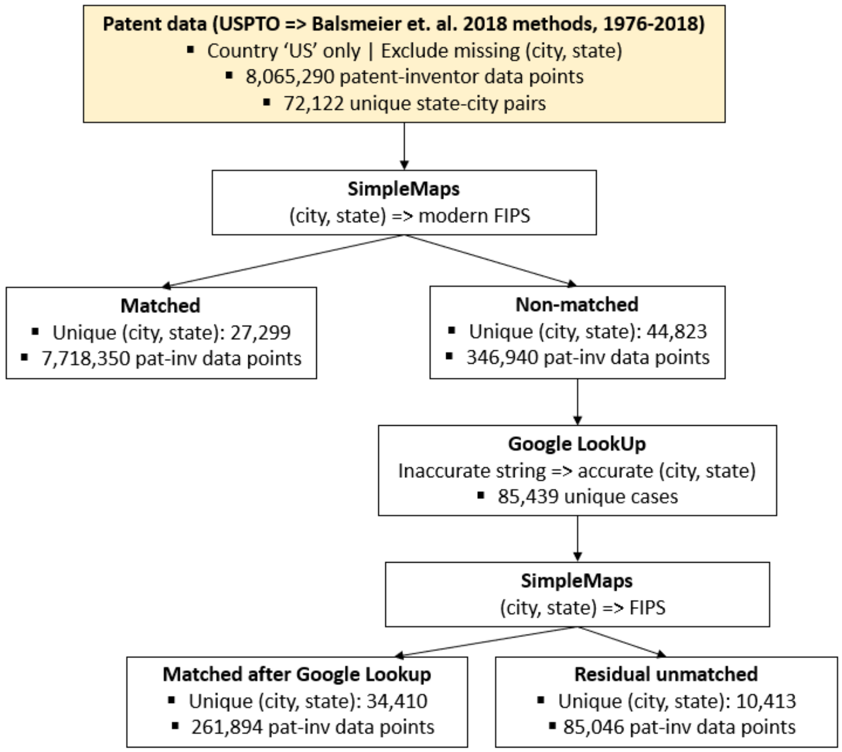


Figure A2: Geographic disambiguation process for U.S. inventor city and state

A3: Concordance between VentureXpert industry groups and NBER patent classification

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. The table details the manual mapping of NBER technological categories to VentureXpert's major industry groups, i.e., Biotechnology, Medical/Health/Life Science, Communications and Media, Computer Related, Semiconductors/Other Electronic, and Non High-tech Technology. As underlying technologies overlap substantially 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, we were able to assign 602,971 inventors to each of the five VentureXpert industry groups.

Table A3: VentureXpert industry groups and NBER patent classification concordance

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

Appendix B: Data description

B1: Spatial distribution of the surname "Marx" in U.S. in 1940

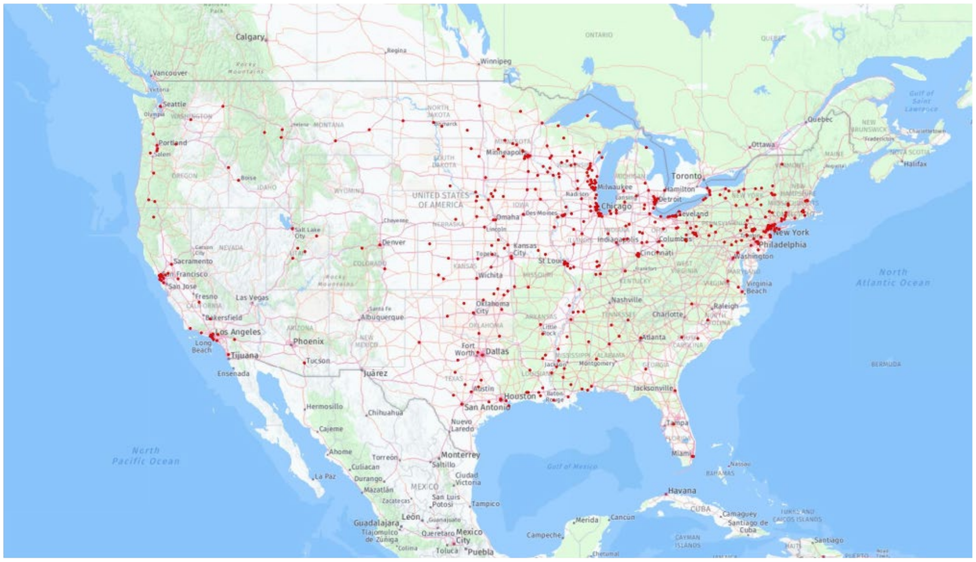


Figure B1: Spatial distribution of the surname "Marx" in 1940 (each red dot = 50 individuals).

B2: Frequency of mobile inventors within the U.S. named Fleming, 1976-2015

Figure B2 illustrates how the number of mobile inventors with surname Fleming varies over time yet does not exhibit a trend (75 moves in total).

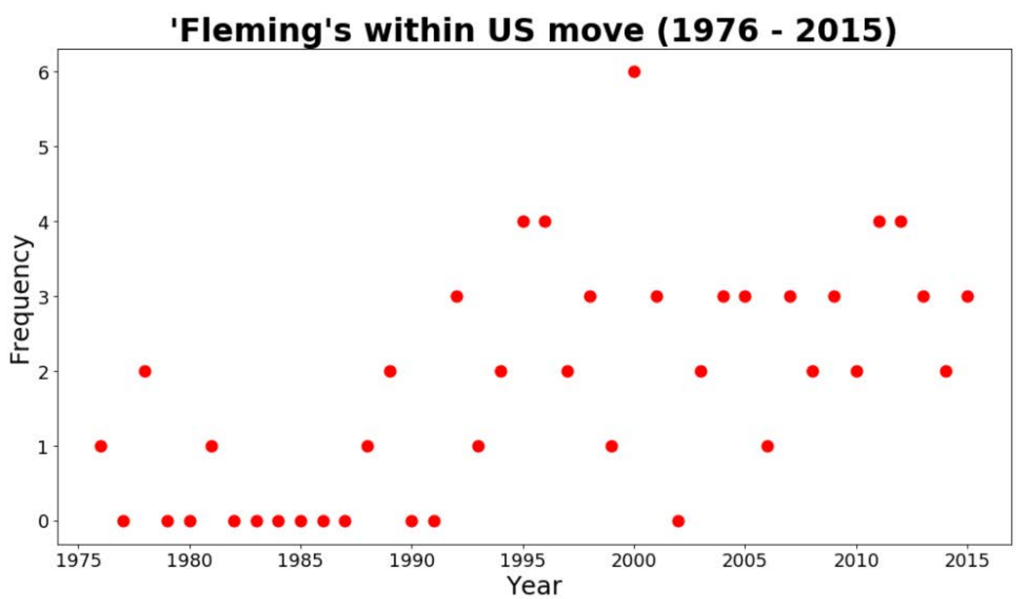


Figure B2: Frequency of mobile inventors named Fleming, 1976-2015.

B3: Destination counties of mobile inventors within the U.S. named Fleming, for the 1980s, 1990s, and 2000s

The maps in Figure B3 show to which counties Flemings moved in the 1980s, 1990s, and 2000s, and the maps in Figure B4 show from which counties Flemings moved in the 1980s, 1990s, and 2000s, illustrating significant variation in origin and destination counties over time. This provides an anecdotal example of why a county-level instrument should be less susceptible than a country-level instrument to the "persistence problem."

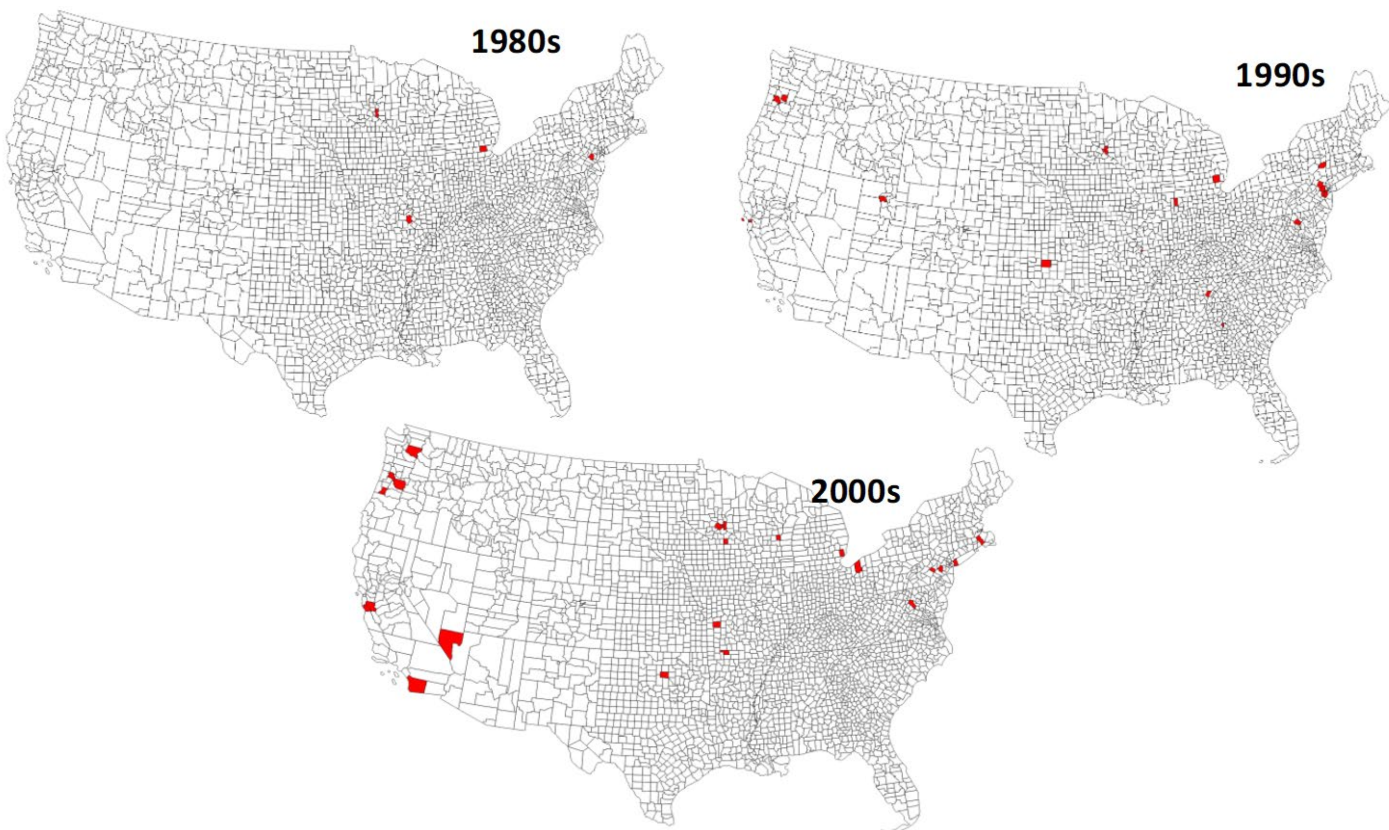


Figure B3: Destination counties of mobile inventors within the U.S. named Fleming, for the 1980s, 1990s, and 2000s.

B4: Origin counties of mobile inventors within the U.S. named Fleming, 1980s, 1990s, and 2000s

Figure B4 illustrates the origin counties from which Flemings moved away in the 1980s, 1990s, and 2000s. It would appear that only one or two of the counties from which Flemings emigrated in the 1990s were also a source of Flemings in the 2000s.

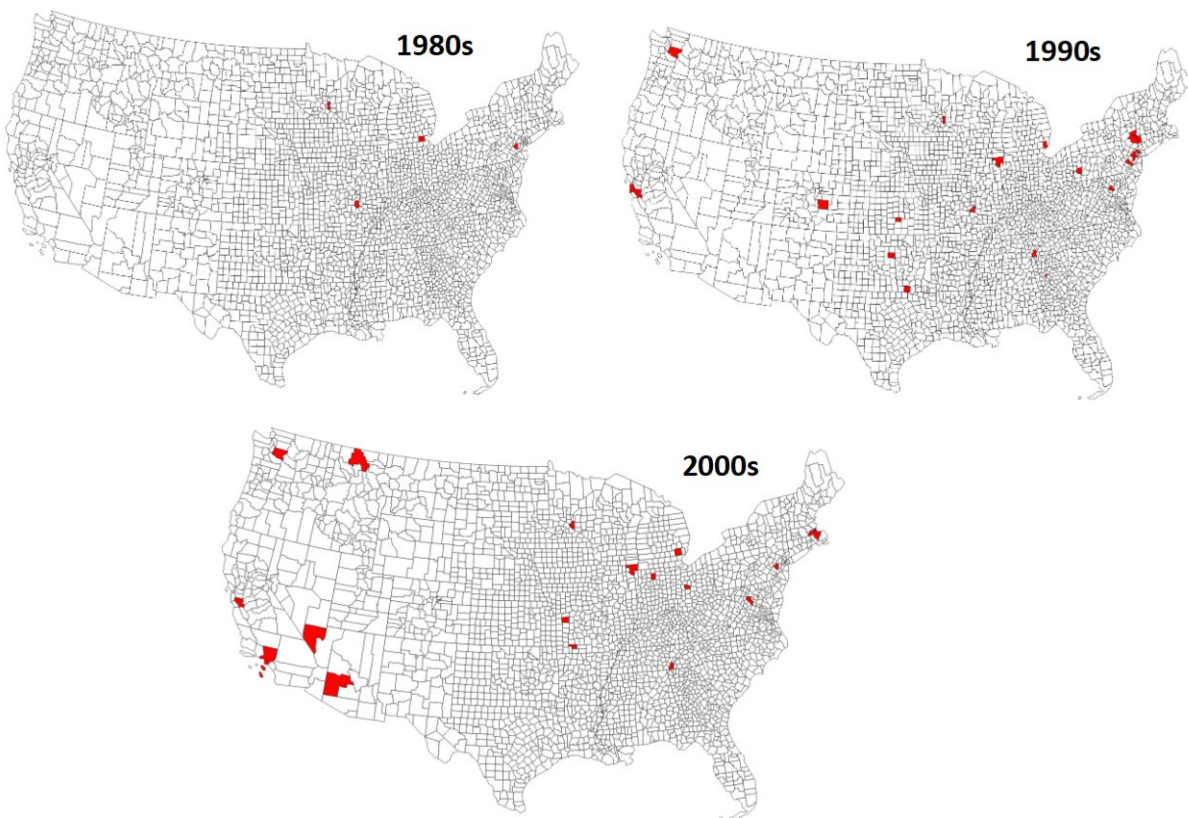


Figure B4: Origin counties of moving inventors within the U.S. named Fleming, for the 1980s, 1990s, and 2000s.

B5: Binned scatter plot of raw inventor mobility and county startup data, 1987-2007

Figure B5 provides a binned scatter plot that illustrates a positive relationship between incoming inventors and new startups in a county.

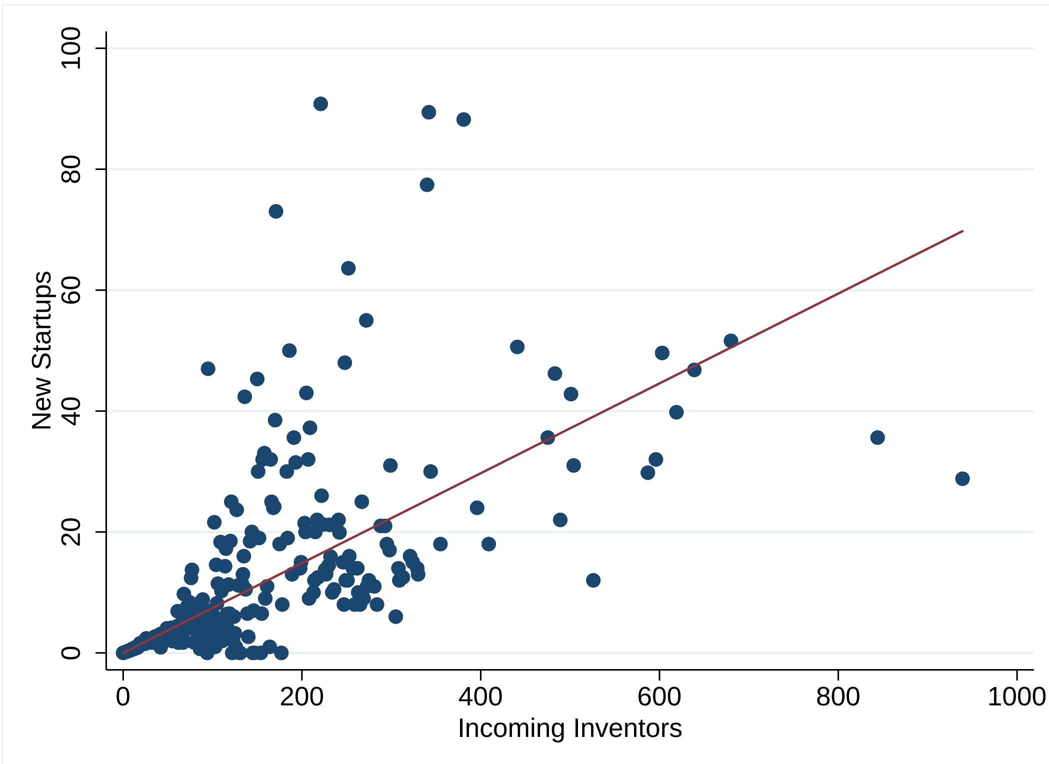


Figure B5: Binned scatter plot of raw inventor mobility and county startup data, 1987-2007.

B6: Yearly count of U.S. inventor mobility by technology field, 1987-2007

Figure B6 breaks out the yearly mobility of US inventors by field. All fields increase almost monotonically, with the exception of non high-tech.

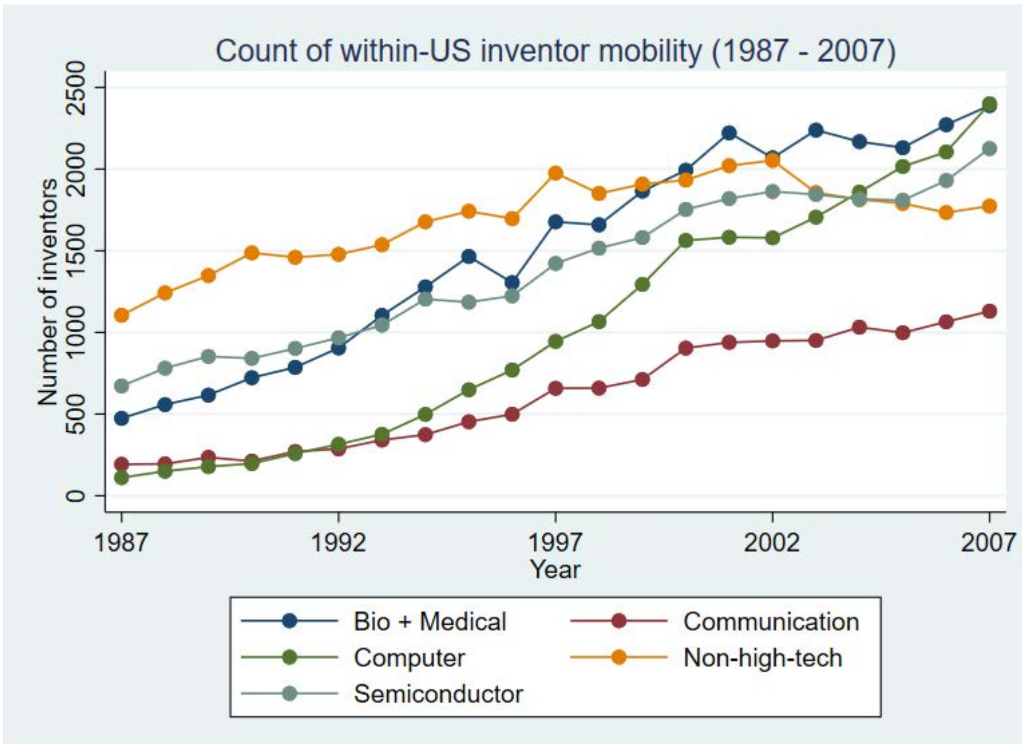


Figure B6: Yearly count of U.S. inventor mobility across U.S. counties by VentureExpert defined technology field, 1987-2007.

B7: Geographical clustering of inventor moves, startups, and successful startups, 1987 to 2007

Figure B7 illustrates the spatial correlation of entrepreneurship and mobile inventors.

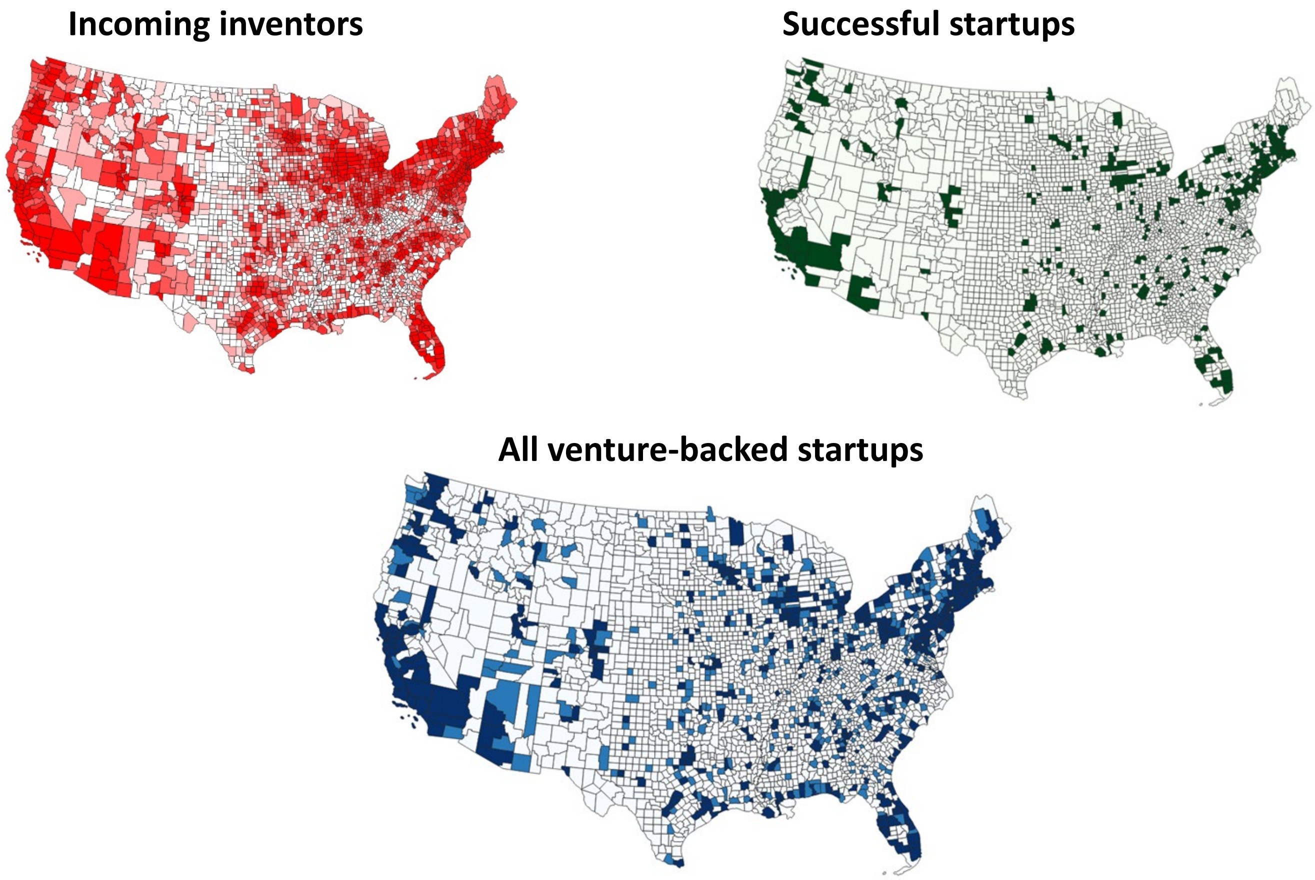
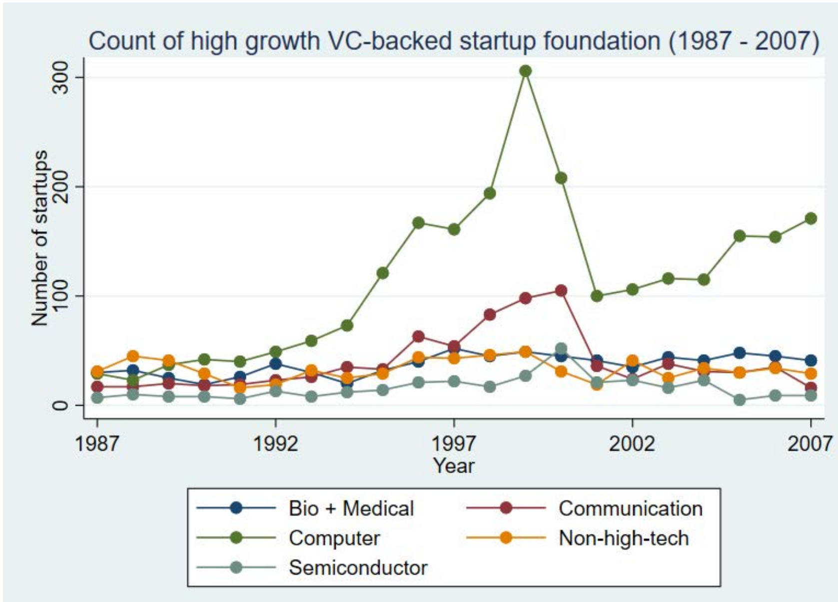


Figure B7: Geographical clustering of inventor moves, startups, and successful startups, 1987 to 2007.

B8: Yearly venture-backed startup creation, in the U.S., 1987 to 2007

Figure B8 illustrates the temporal trends on VC backed startups, by field.



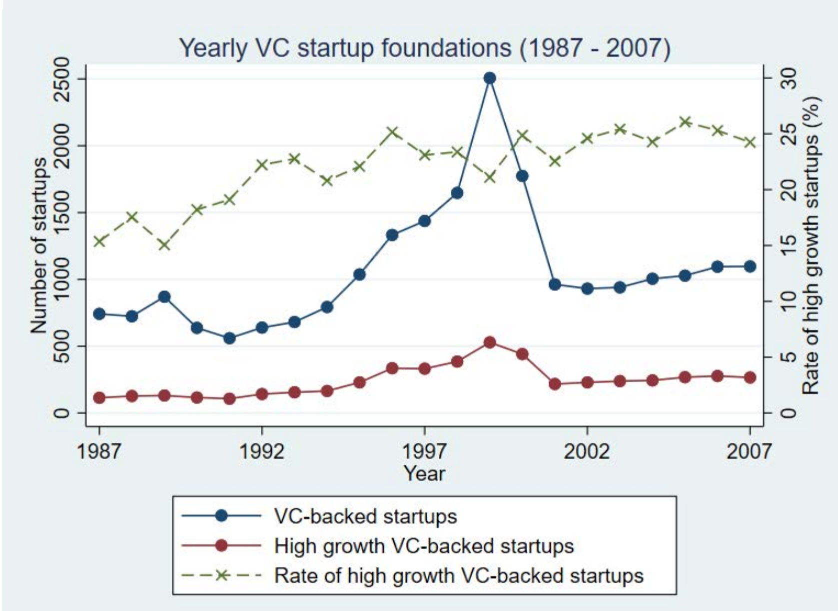


Figure B8: Yearly venture-backed startup creation, in the U.S., 1987 to 2007

Appendix C: First stage plausibility check

C1: First stage regressions of historical surname distribution and modern mobility

We establish the plausibility of the instrument's first stage by investigating the linkage between the historical surname distribution and the geographical mobility of individual inventors. This approach builds on an extensive demographic literature, including studying the migration of people, social networks and mobility (Rossi, 2013), tracking migration and mobility using surname distribution in Italy and France (Piazza et al. 1987, Darlu et. al. 2011), inferring the geographical origin of migrants with surnames (Degioanni & Darlu 2001), using surnames to estimate academic mobility (Grilli & Allesina 2017) and social mobility and intergenerational wealth transfer (Clark & Cummins, 2014, 2015).

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.[5] We empirically 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 realized and each potential destination county, include the share of people in the 1940 Census with the same surname. Using this dyadic dataset covering 258,657 moves from 1988-2014, we estimate the following model with OLS:

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

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.

Table C1 presents estimations for 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.

[5] 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. 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.

Table C1 – First stage validation check: 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.

All specifications consistently show that an increase in the historic surname share in a potential destination county correlates with a significantly higher probability of observing a given inventor moving to that specific destination county as compared to all other potential destination choices. Note that since the dependent variable vector is sparse a low R^2 is to be expected. The large increase in explained variation when destination and destination-origin county fixed effects are included illustrates how unobserved time invariant factors also explain mobility decisions.

Appendix D: Shift share instrumental variable (SSIV) validation

Due to the novelty of the instrument and recent innovations in the econometric literature (Adao et.al. 2019; Goldsmith-Pinkham 2020; Borusyak et. al. 2022), we present a variety of analyses to explore the validity of the instrument. While recent methodological research agrees that causality is sufficiently established if either the share or the shift is conditionally exogenous (Goldsmith-Pinkham et al. 2020; Borusyak et al. 2022), we explore the validity of our approach under either set of assumptions.

It should also be noted that both stages of the IV include county and time 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. As in the canonical difference-in-differences setup, we only exploit variation from the interaction of a share varying at the unit level, i.e., the proportion of historic surnames in a county, and shocks varying at the time dimension outside of that county. Variation in our instrument should thus not impacted by differences in levels of any unobserved county characteristic. The key assumption we will explore is that, conditional on unit and time fixed effects, the instrument remains exogenous and can thus be used to estimate the causal impact of inventor inflows on startup creation using Equation (1).

We group and first discuss the share, and then the shift, component of the instrument.

D1: Share instrument remains uncorrelated with modern pull factors

Inventors are probably attracted to wealthy, large, and economically vibrant counties. Indeed, as shown in model a, inbound mobility strongly correlates with these factors. The instrument, however, does not, as shown in model b.

Table D1: Regression of incoming inventors and shift-share instrument on modern observable pull factors

	a	b
	Incoming Inventors$_{t-1}$	Shift-share instrument$_{t-1}$
Income$_{t-1}$	0.505***	-0.007
	(0.042)	(0.043)
N	64,133	64,133
within-R^2	0.021	0.000
Employment$_{t-1}$	0.324***	-0.030
	(0.038)	(0.030)
N	52,788	52,788
within-R^2	0.006	0.001
Population$_{t-1}$	0.694***	-0.021
	(0.053)	(0.065)
N	64,133	64,133
within-R^2	0.024	0.000

Notes: This table presents OLS regression of treatment variation (i.e., incoming inventors) and the shift-share instrument, column (a) and (b) respectively, "on observable contemporary factors," including yearly income level and population of counties obtained from the Bureau of Economic Analysis as well as yearly employment of counties obtained from the U.S. Bureau of Labor Statistics. Each model includes state-year and county fixed effects. Incoming inventors, the instrument as well as the observable factors are log-transformed. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively. Note that the number of observations varies according to the availability of values on regional factors.

D2: Share instrument remains uncorrelated with most historical regional characteristics and future change in startup founding rates

One possible concern is that county-level characteristics in 1940 predict modern entrepreneurial activity, calling into question the exogeneity of the shares, despite the Fixed Effects estimations. To explore this possibility, we begin by regressing our dependent variable -- the yearly change in modern entrepreneurship -- on historical county characteristics. First, we test the gender ratio of the county, given extensive evidence of a gender gap in entrepreneurship (Guzman & Kacperczyk, 2019; Miric & Yin, 2000; Marx, 2021). Second, we test the racial composition of counties given studies suggesting bias against minority entrepreneurs (Chatterji & Seamans, 2012; Younkin & Kuppuswamy, 2018) and recent evidence of dramatic under-representation of racial minorities among venture-backed ventures (Wang & Marx, 2022). Third, we test educational levels in a county given that venture investors are less likely to fund startups in low-skilled sectors, in favor of sectors that require higher levels of human capital and education. Finally, we test average historical income levels (to measure changes, the dependent variable is de-meaned by the county's average level of entrepreneurship).

Precisely estimated coefficients for any of these three factors would raise the concern that modern entrepreneurial activity, including the founding of successful venture-backed startups, is strongly tied to historical regional characteristics. While Table D2 illustrates no relationships between the change in the number of successful modern startups founded in the focal county and the 1940 ratio of male citizens, the ratio of white citizens, the ratio of citizens who completed 12th grade or college, it does illustrate a correlation with average income level. Hence, in the next Appendix, we dropped correlated surnames in instrument construction.

Table D2: Correlation between historical regional characteristics and change in modern outcome variable of entrepreneurship

	Successful venture-backed startups founded	Obs.
Ratio of Male	-4.88E-9 (3.73E-9)	64,890
Ratio of White	1.31E-10 (2.03E-10)	64,890
12th grade or College	3.36E-9 (2.16E-9)	64,365
Average income	1.71E-12** (6.71E-13)	64,365

Notes: This table presents OLS regression of log(number of successful venture-backed startups founded + 1) demeaned within each county. "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. Each model includes state-year fixed effects. Standard errors clustered at the state level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

D3: Share instrument remains robust to removal of correlated surnames

Because we found in the regression above that historical county-level average income correlates with changes in modern entrepreneurial activity, we show that our results hold when the instrument excludes all surnames which are correlated with a modern county characteristic. To do this, we first run a regression for every individual surname in order to establish its correlation with the historical county characteristics. We then show that the purged instrument remains relevant for predicting local inventor inflows and that our final estimates remain stable in terms of statistical significance as well as economic magnitude.

For each surname, we ran an individual regression of historic surname fraction on a historical regional characteristic from the 1940 Census (Tables a-c). We then plotted p-value vs. coefficients for each characteristic and noted the percentage and data points of significant regressions with $p < 0.05$ (to the left of the red line in plots). This exercise supports the assumption that there is no unobserved factor correlated with historic surname shares that would simultaneously predict inventors moving to a given county and changes in modern entrepreneurial activity. Note that even before these analyses, that differences in *levels* of historic county characteristics, surname characteristics and startup rates should not pose a threat to identification, as those should be absorbed by unit fixed effects.

Table D3a: Ratio of Male: significantly correlated with 6.9% of surnames' shares

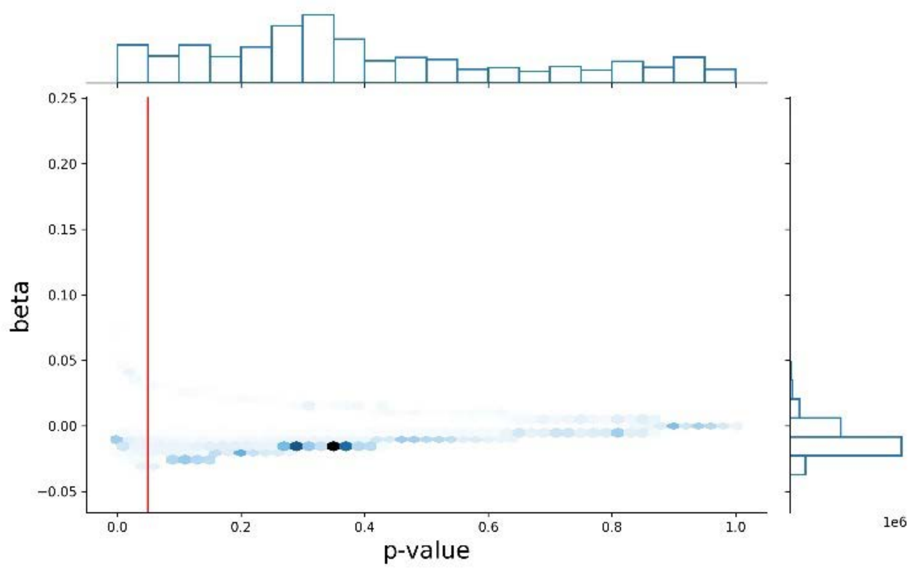


Table D3b: Ratio of White: significantly correlated with 2.6% of surnames' shares

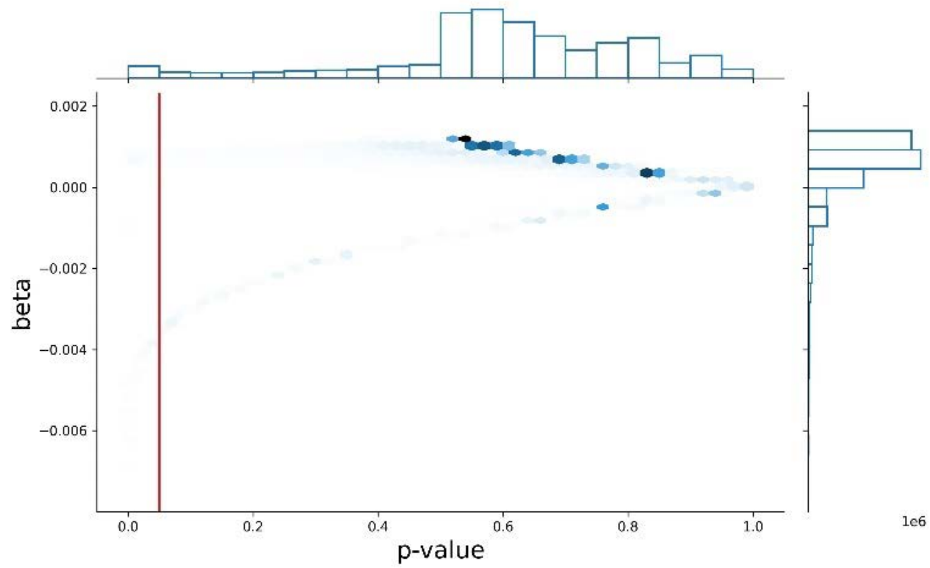


Table D3c: Level of education (Ratio of 12th grade or higher education): significantly correlated with 17.0% of surnames' shares

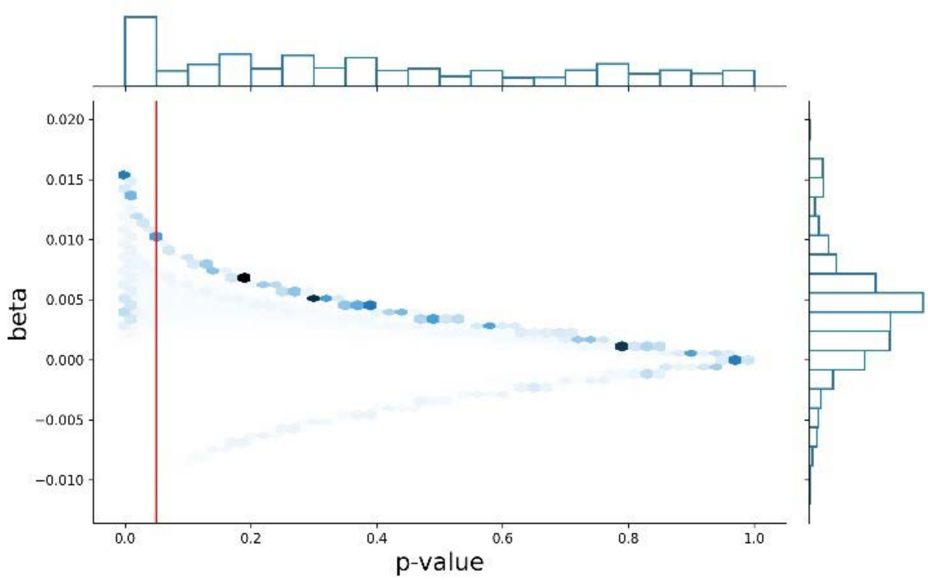


Table D3d: Average income: significantly correlated with 53.5% of surnames' shares

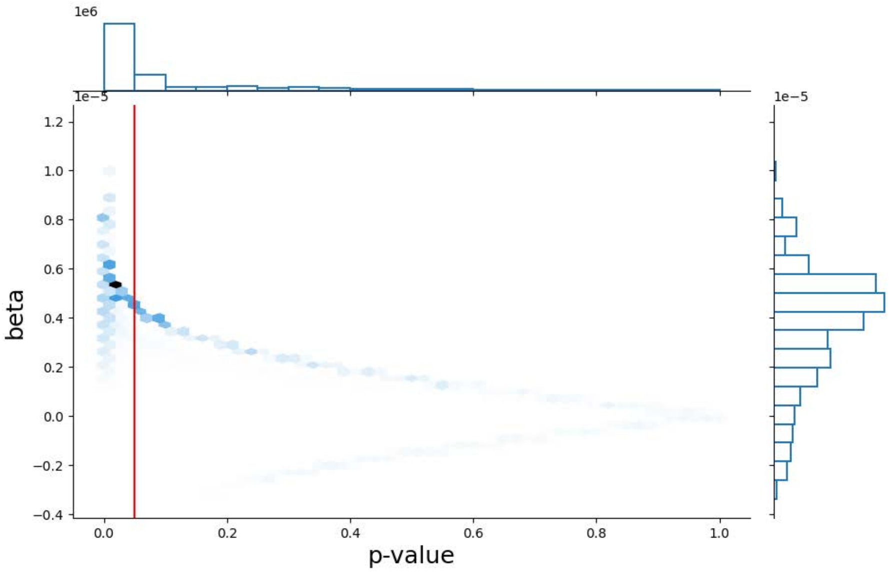


Table D3e re-estimates our regressions while excluding any name where the individual regression was significant at $p < 0.05$, i.e., all points on prior graphs to left of the red line.

Table D3e: Results remain robust to removal of correlated surnames from the instrument

	a	b	c	d
	Exclude surnames correlated w/ ratio of male	Exclude surnames correlated w/ ratio of white	Exclude surnames correlated w/ ratio of educated to non-educated (12th grade or college)	Exclude surnames correlated w/ average income
	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	0.159***	0.122***	0.149**	0.302**
	(0.053)	(0.038)	(0.071)	(0.137)
N	65,247	65,247	65,247	65,247
First Stage F	95.049	150.073	55.777	18.795
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) excludes surnames that are correlated with the ratio of male; (b) excludes surnames that are correlated with the ratio of white; (c) excludes surnames that are correlated with the education level (measured by the fraction of people received 12th grade or college education; (d) excludes surnames that are correlated with average county income. 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.

D4: Share instrument individual treatments are homogenous

Although Appendix C established the plausibility of the first stage, we further investigate that the first- and second-stage results do not mask heterogeneity in individual treatment effects. Similar to recently articulated concerns for difference-in-differences models with multiple continuous treatments (de Chaisemartin and D'Haultfoeuille 2020; Goodman-Bacon 2021), caution would be warranted if an analysis of individual treatment effects revealed a substantial number of negative relationships with inventor mobility, because we should expect only positive effects (negative effects would also imply that families prefer to live away from one another, which would be inconsistent with prior research, i.e., Darlu et. al. 2011 and Clark & Cummins 2015).

To investigate potential heterogeneity with respect to surname shares that correlate with incoming inventors of the same name, we ran individual regressions of mobility into a particular county on the historical surname fraction for the same name in that county (e.g., the number of Balsmeiers moving into a given county in a year is regressed on the 1940 share of Balsmeiers in that county). By construction, this regression is restricted to individual inventor surnames that moved in the sampling period (33,444 different surnames). Figures D4a and D4b illustrate the results of these 33,444 regressions. We would be concerned if the predicted coefficients were a heterogeneous mix of positive and negative coefficients, suggesting that even if most surname shares predicted inbound inventor mobility, other surname shares predicted the opposite. If this were true, we would expect to see a large number of negative and precisely-estimated dots in the lower-left-hand section of Figure 4Db, where we zoom into the 90% of data that does not involve outlier predictions.

As is visible in Figure 4Db, there do not appear to be any precisely estimated negative predictions (i.e., to the left of the red line and below zero). In fact, there are only a handful of negative predictions, and almost all are rather imprecisely estimated ($p>0.5$). The vast majority of predictions have a positive sign, and although a number of these are also imprecisely estimated, we find ~20% of surnames to be individually significant ($p < 0.05$). This might appear low but should not be entirely surprising given the baseline rarity of mobility events, as well as the fact that moves occur for many reasons besides the location of the inventor's extended family (e.g., partner's family, climate, cost of real estate, schools for children, etc.). Unlike the negative and significant correlations demonstrated in Goldsmith-Pinkham et al. (2020), nearly all of our correlations demonstrate a positive coefficient, suggesting homogeneity in individual treatment effects.

Table D4a: Regressions of mobility on historical surname fraction in county for mobile surnames

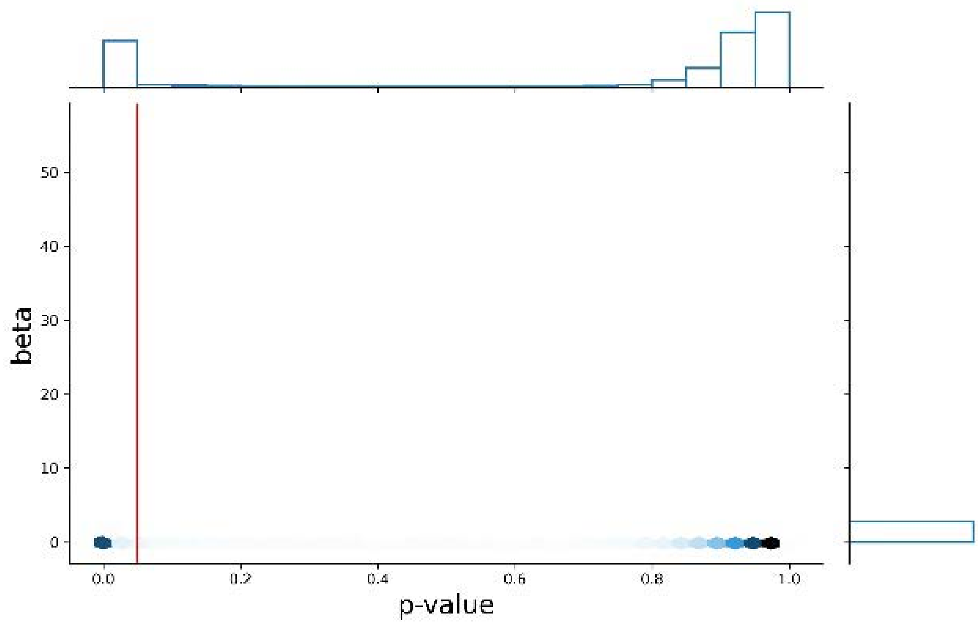
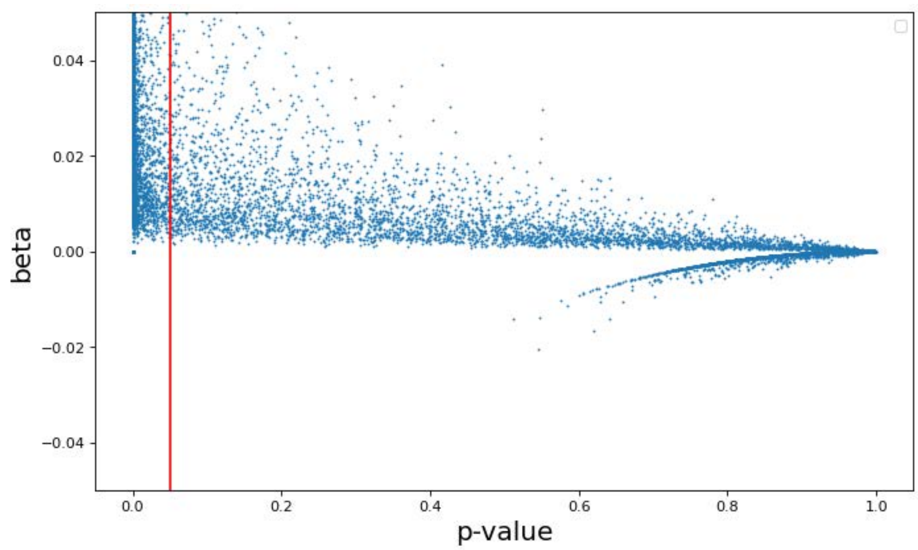


Table D4b: Closeup of regressions of mobility on historical surname fraction in county for mobile surnames



Note: Table D4b zooms in to show the distribution of plots in detail, highlighting betas between -0.05 and 0.05. (Note that 90.8% of data points are included in this plot), and in particular, all negatively estimated coefficients).

Armed with these measures of the predictive power for each individual surname (i.e., the beta coefficient of each individual regression), we additionally tested whether the instrument might be vulnerable to an ethnicity bias. One might imagine that moving to a particular county is driven more by ethnicity than by family ties (as proxied via shared surnames). Using information from the 1940 Census we first construct dummy variables indicating family names with a majority of people being labeled as Black, Asian, American Indian, or White, respectively. We then regress the predicted correlation for each of the 33,444 surnames from Figure D4a on these ethnic indicators (in a single regression).

Even though we do not have indicators for all possible ethnicities, if inbound mobility in a county were driven substantially by such factors, we might expect at least one of these ethnic indicators to have a positive and precisely estimated coefficient. However, Appendix Table D4c also shows that none of these indicators has a significant correlation with the size of beta, i.e., the strength with which a given surname predicts a move of an inventor with the same surname. In fact, most estimated coefficients are negative (albeit imprecisely estimated). This remains true whether evaluating all surnames (model a) or only those with statistically significant betas (model b, i.e., those to the left of the red line in Table D4b).

Although it is impossible to formally prove the exogeneity of the instrument shares, the stability of the results after various substantial adjustments suggests that the high number of historic surname shares and destination counties contribute to a plausible SSIV -- and arguably meet the conditions for a causal interpretation based on the exogeneity of shares assumption (Goldsmith-Pinkham et al. 2020).

Table D4c: Heterogeneity in the marginal effect of surname share and lack of ethnicity bias

	Marginal effect of surname share (β)	
	a	b
	All surnames ever moved in the sampling period	Only surnames with significant share effect
Ethnicity		
White == 1	0.017 (0.024)	0.098 (0.109)
Black == 1	-0.039 (0.043)	-0.168 (0.296)
American Indian == 1	-0.042 (0.146)	(No names with significant share effect)
Asian == 1	-0.005 (0.030)	-0.087 (0.117)
N	33,444	6,822

Notes: This table presents OLS regression of marginal effect sizes of surname shares on the heterogeneity of individual surnames, i.e., ethnicity. While column (a) shows the results based on a sample with all surnames ever moved in the sampling period, column (b) shows the results based on a sample with only surnames whose historic surname fractions have a significant marginal effect on the inventor mobility. Ethnicity variables are indicator variables. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

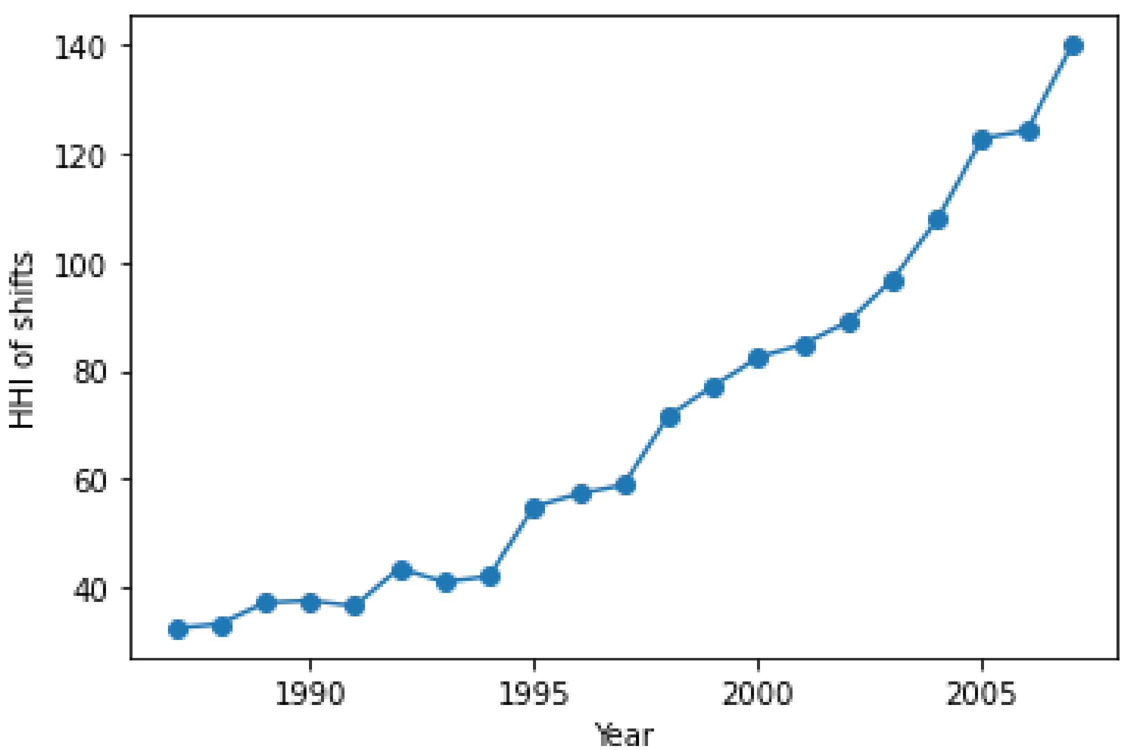
D5: Shift instrument demonstrates low sum of HHI measures

Borusyak et al. (2022) show that SSIVs can be valid even in the absence of share exogeneity if shocks are large in number and as good as randomly assigned. Although appendices D1-D4 suggest that our SSIV may be valid with regards share-endogeneity concerns, we also analyze the validity of the shift (and follow the convention of using 'shift' and 'shock' interchangeably).

Intuitively, the number of moving inventors with the name 'Fleming' in a given year should not correlate with any single county's number of startup foundings – especially if the 'Flemings' moving into the specific county are purged from the shock, as is the case with our leave-out SSIV construction in equation (3). Further, given that we observe more than 200,000 inventor surnames, it seems unlikely that our results are driven by any single shock or group of shocks.

Borusyak et al. (2022) formalize this intuition with the assertion that the sum of the quadratic shock proportions $\sum_{n,t} s_{nt}^2$ should be minimized and ideally approach zero and that the inverse of the Hirschman Herfindahl Index (1/HHI = $1/\sum_{n,t} s_{nt}^2$) is a good metric with which to proxy the effective sample size that drives identification. For our baseline sample, Table D5 shows the HHI to vary between 0.00324 (sample year 1987) and 0.01401 (sample year 2007), which translates into effective sample sizes of 308.1 and 71.38.

Table D5: HHI of shifts per year



Note: HHI values here are multiplied by 10,000

- Min: 32.457 (in 1987); Max: 140.116 (in 2007)

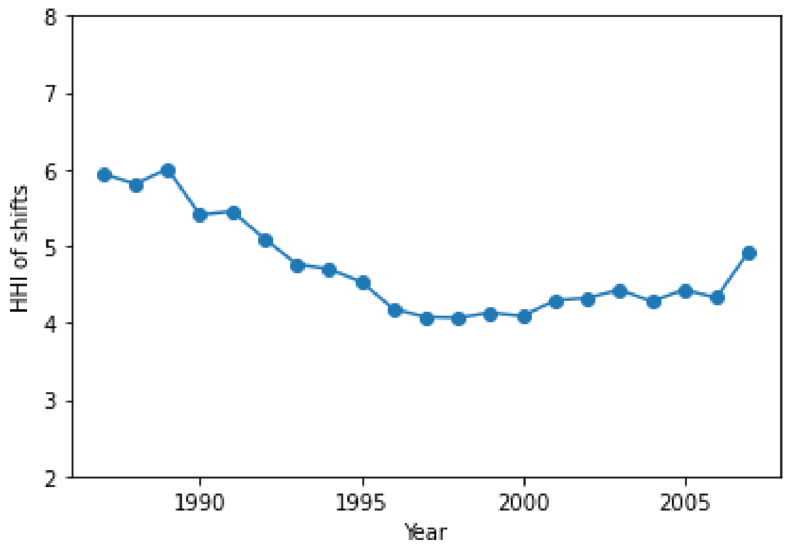
D6: Shift instrument demonstrates low sum of HHI measures, excluding top 50 most popular surnames

Since the distribution of shocks is skewed towards the most frequent surnames, we also re-ran this validation check without the top 50 most frequent surnames (consistent with the alternate instrument that proved robust in section 5.1.5). Table D6a shows the list of the top 50 most popular shift surnames between 1987 and 2007. In Table D6b we find the HHI to vary between 0.00041 (sample year 1998) and 0.00060 (sample year (1989), which translates into effective sample sizes ranging from 2439.02 to 1666.07. Note that these numbers are large (and preferred) in comparison to prominent SSIV results in the literature, e.g. Autor et al. (2013) who report sample sizes between 1.7 and 191.6.

Table D6a: Top 50 most popular shift surnames between 1987 and 2007

Rank	Surname	Mobility frequency	Rank	Surname	Mobility frequency
1	smith	2719	26	white	355
2	johnson	2098	27	huang	354
3	miller	1801	28	king	350
4	wang	1144	29	young	344
5	brown	1138	30	allen	343
6	chen	1106	31	peterson	342
7	anderson	1042	32	lewis	321
8	lee	1011	33	kim	316
9	williams	1006	34	harris	303
10	jones	885	35	chang	298
11	li	785	36	adams	296
12	davis	764	37	baker	272
13	wilson	658	38	walker	271
14	liu	632	39	wright	270
15	martin	605	40	xu	260
16	nelson	489	41	yu	235
17	taylor	487	42	jackson	232
18	thompson	480	43	wong	230
19	clark	426	44	kelly	229
20	wu	422	45	green	224
21	moore	411	46	mitchell	219
22	yang	404	47	roberts	218
23	hall	380	48	evans	218
24	thomas	368	49	lu	218
25	lin	361	50	edwards	217

Table D6b: HHI of shifts per year excluding top 50 most popular surnames



Note: HHI values here are multiplied by 10,000. Min: 4.066 (in 1998); Max: 6.006 (in 1989)

D7: Shift instrument demonstrates low serial correlation

Assuming shocks to be as good as randomly assigned might be problematic if they remain influenced by unobserved trends. While a graphical inspection of the 'Fleming' shocks presented in the Appendix Tables B2-4 illustrates no clear trend, we also perform a more systematic test for serial correlations and trends by running individual regressions for every individual surname observation on the prior period's individual observation (individual regressions of the shift in year *t* on the shift in *t*-1, for each surname), as well as a linear trend. Figures D7a-b illustrate that 5.78% of surname measures are serially correlated and Figures D7c-d show that 5.80% of surname shocks show a significant linear trend. Both of these numbers are close to a purely random association of 5%.

Nonetheless, in Table D7e, we show that our results are not influenced even if we exclude those potentially problematic shocks from the instrument construction. Model a excludes the 5.78% of surnames that were identified as exhibiting serial correlation. The 5.80% of surnames exhibiting linear time trends are dropped from model b. Both exclusions yield little impact on the results.

Figure D7a: Serial correlations of shocks, for each surname

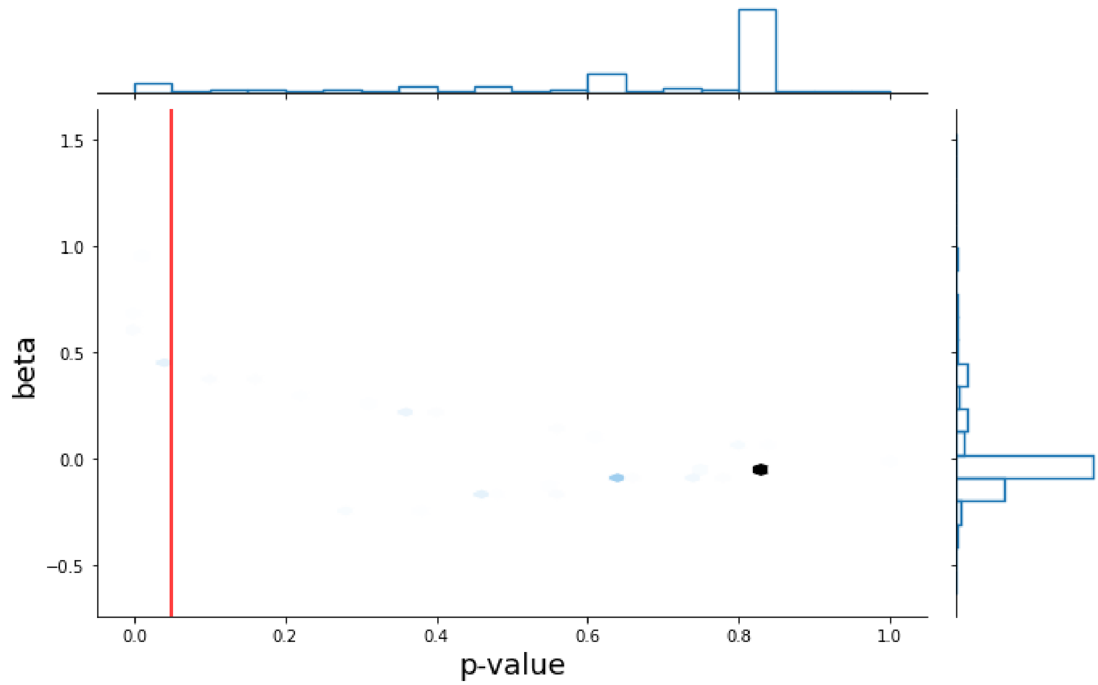


Figure D7b: Serial correlations of shocks, for each surname, scatter plot

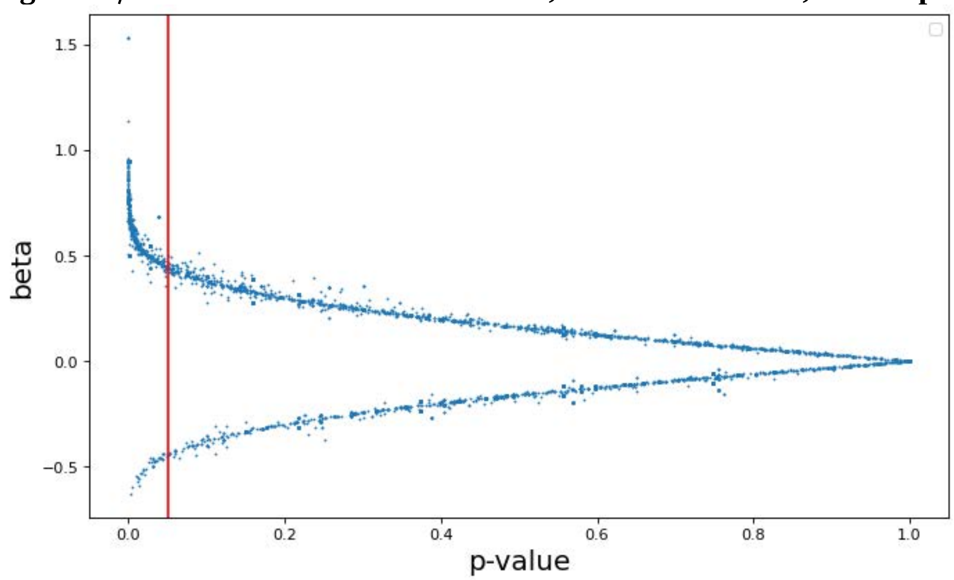


Figure D7c: Serial correlations of shocks and a linear trend, for each surname

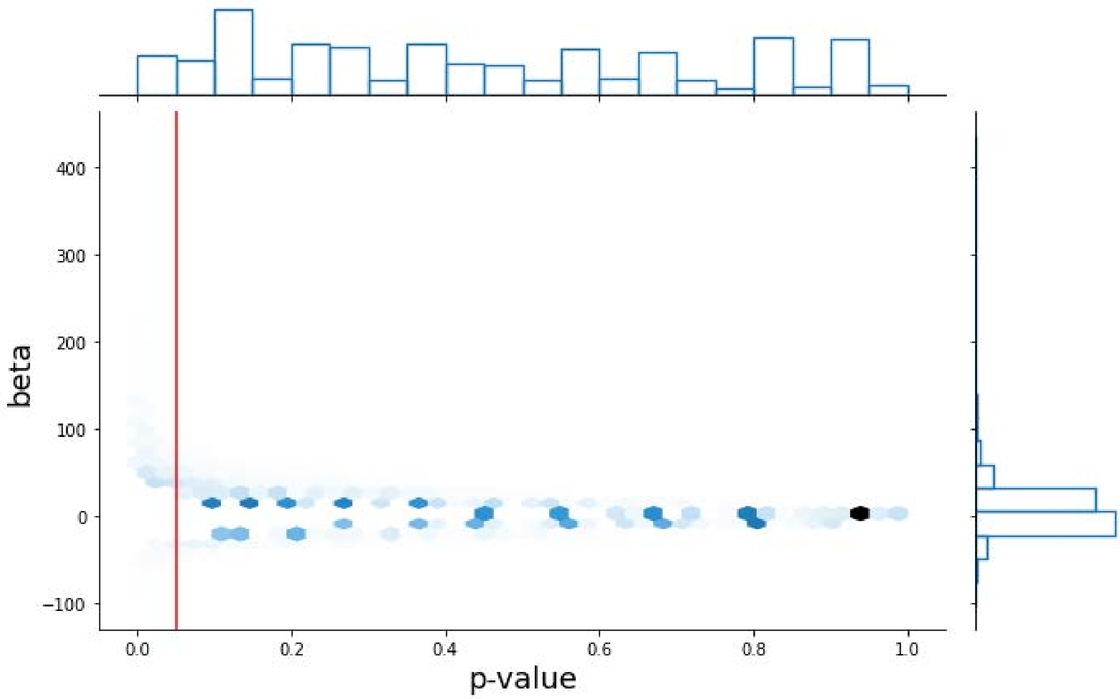


Figure D7d: Serial correlations of shocks and a linear trend, for each surname, scatter plot

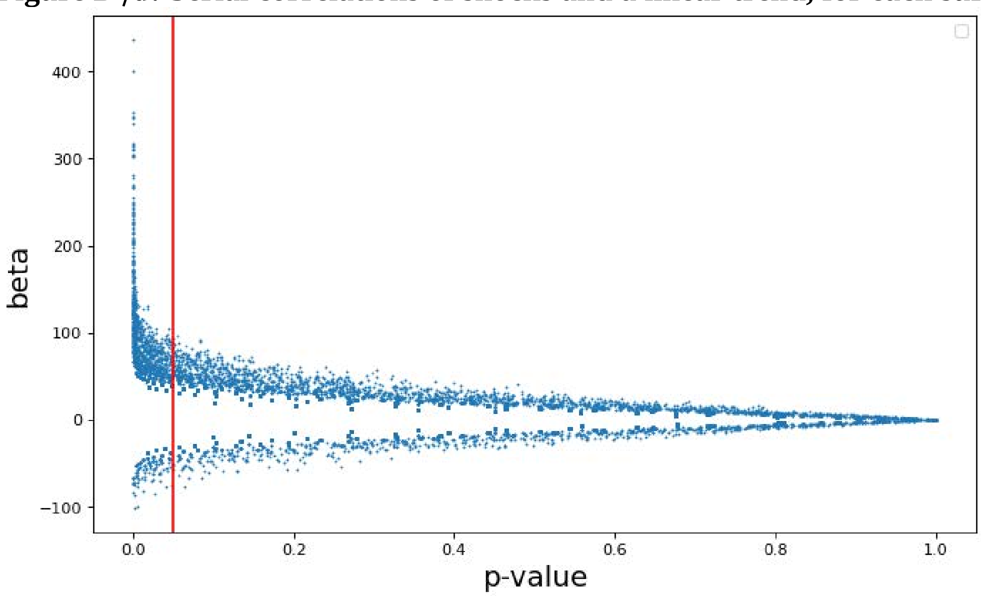


Table D7e – Alternative instruments which remove inventors with surnames whose mobility is serially correlated or exhibit significant time trends

	Successful venture-backed startups founded	
	a	b
	Exclude surnames from shift with serially correlated mobility	Exclude surnames from shift that have significant time trends in mobility
	IV	IV
Incoming Inventors$_{t-1}$	0.095***	0.107***
	(0.028)	(0.028)
N	65,247	65,247
First Stage F	223.699	221.343
State-Year FE	Yes	Yes
County FE	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) excludes surnames with serially correlated shifts; (b) excludes surnames with shifts that have significant time trends. 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.

D8: Shift instrument demonstrates robustness to unobserved spatial correlations

To rule out potentially problematic spatial correlations of shocks, we provide Table D8, showing that our SSIV produces similar results if we leave out not only same-destination-county moves (as in equation (3)) but also same-state moves (that is, all counties of the state the county is in) in model c. This expanded leave-out approach helps to allay concerns that moves to nearby counties, but not the focal county, might influence entrepreneurial activity in the focal county. Table D8 also illustrates robustness to using all U.S. inventors as the shift, independent of whether they moved or not (model a) or even when constructing the shift only from inventors who did *not* move (in which case the leave-out adjustment is not necessary).

Overall, these additional shift tests support the assumption that our surname specific shocks are large in number and hopefully as good as randomly assigned. The SSIV could thus arguably remain valid even in the absence of conditional share exogeneity (Borusyak et al. 2022).

Table D8: Alternative shift instruments

	Successful venture-backed startups founded		
	a	b	c
	Total inventors in US	Non-moving inventors	State leave-out
	IV	IV	IV
Incoming Inventors$_{t-1}$	0.148**	0.143***	0.103***
	(0.062)	(0.060)	(0.036)
N	65,247	65,247	65,247
First Stage F	24.413	25.717	183.772
State-Year FE	Yes	Yes	Yes
County FE	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) uses the count of total inventors in US at the national level as the shift; (b) uses the count of non-moving inventors at the national level as the shift; (c) leaves out states' own inflows, instead of counties' own inflows, from national flows with the same 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.

Column (c) of Table D8 is helpful in understanding the limitations of the SSIV, especially regarding demand shocks. To fix ideas, consider a scenario where universities in Buffalo NY, Albany NY, and Atlanta GA achieve unprecedented breakthroughs in EV battery technology during 2014. These demand shocks lead to the founding of VC-backed EV startups the following year in those counties. Assume also that inventors named Balsmeier are particularly skilled in EV technology and therefore moved to Buffalo, Albany, and Atlanta in 2014, increasing that year's nationwide "shift" for Balsmeier. Finally, assume that in 1940 there were some Balsmeiers in Albany and Buffalo but none in Atlanta.

The leave-out provision of the instrument prevents the additional Balsmeiers moving into Buffalo during 2014 (due to the EV shock) from being counted toward the calculation of the instrument. Therefore, a demand shock that occurs only in one county should not threaten the instrument. However, the additional Balsmeiers moving into Albany would still count toward Buffalo's instrument; therefore, a simultaneous demand shock might threaten the instrument. However, column (c) of Table C5 above shows that the instrument is also strong when the leave-out provision is calculated at the state level, where the additional Balsmeiers moving into Albany do not count toward Buffalo's instrument. Therefore this concern is limited to demand shocks that occur simultaneously in multiple states. We set aside Albany for the remainder of our example.

In our example, the additional Balsmeiers moving into Atlanta would count toward Buffalo's instrument. But because Atlanta did not have any Balsmeiers in 1940, the additional Balsmeiers moving into Buffalo do not count toward Atlanta's instrument. Even for Buffalo, it is not clear how much the additional Balsmeiers moving into Atlanta will affect its instrument if Buffalo had a very small number of Balsmeiers in 1940. We might be more worried if Wilsons were highly skilled at EV technology, although in that case models (b) and (f) of Appendix Table C7 show that the instrument is strong when omitting the top 10% of surnames.

In sum, there is a risk that our instrument can suffer from simultaneous, industry-specific demand shocks in multiple states where the surnames of inventors skilled in that industry who moved to the shocked counties in other states were well represented in 1940 for a focal county.

D9: Main specification re-estimated with local income

To control for potential confounding effects of local income, we tested the same specifications as Table 2 but controlling for local per capita income in period t-1. Per capita income is calculated based on yearly income level and population of counties obtained from the Bureau of Economic Analysis. The results remain consistent.

Table D9: Impact of incoming inventors on local venture backed startups, controlling for local per capita income in period *t*-1

	Incoming Inventors$_{t-1}$	Number of venture-backed startups founded				
	a	b	c	d	e	f
	OLS (first stage)	OLS	OLS	OLS	IV	IV (w/o top 10 counties)
$\widetilde{Inv_{d,t-1,leave-out}}$	0.755***					
	(0.032)					
Incoming Inventors$_{t-1}$		0.341***	0.342***	0.031***	0.179***	0.119***
		(0.019)	(0.019)	(0.005)	(0.039)	(0.035)
Per Capita Income$_{t-1}$	1.411***	0.116***	0.156***	0.078***	0.048***	0.037***
	(0.024)	(0.033)	(0.042)	(0.016)	(0.014)	(0.013)
N	64,133	64,133	64,133	64,133	64,133	63,944
First stage F					183.146	177.432
Year FE	Yes	Yes	No	No	No	No
State-Year FE	No	No	Yes	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes
R^2	0.545	0.488	0.516	0.836		

Notes: This table reports the results of specifications (a) – (f) from Table 2, with controlling for local per capita income in period t-1. Specification (a) in this table present the first stage OLS results of the regression of incoming inventors on the IV (i.e. $\widetilde{Inv_{d,t-1,leave-out}}$). Specification (b) – (d) present OLS regressions of log (number of venture-backed startups + 1). Incoming inventors as well as the instrument are log-transformed. Specification (b) includes year fixed effects. Specification (c) includes state-year fixed effects. Specification (d) includes state-year and county fixed effects. Specifications (e) shows the results of our IV regression with state-year and county fixed effects, where incoming inventors are instrumented with $\widetilde{Inv_{d,t-1,leave-out}}$ in the first stage as defined in equation (3). Specification (f) shows results of our IV regression, but excluding the top 10 entrepreneurial counties from the sample (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). 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. Note that the county-year observations with missing data on the variable controls are dropped in this analysis.

D10: Main specification re-estimated without the right tail of the SSIV distribution

One may be concerned that the result may be driven by the right tail of the distribution of SSIV. To investigate this, we excluded the top 5% of observations with the highest values of SSIV and modeled the same specifications as Table 2. The results remain consistent after excluding these observations.

Table D10: Impact of incoming inventors on local venture backed startups, excluding the top 5% of observations with the highest values of the instrumental variable

	Incoming Inventors$_{t-1}$	Number of venture-backed startups founded				
	a	b	c	d	e	f
	OLS (first stage)	OLS	OLS	OLS	IV	IV (w/o top 10 counties)
$\widetilde{Inv_{d,t-1,leave-out}}$	0.755***					
	(0.032)					
Incoming Inventors$_{t-1}$		0.227***	0.232***	0.018***	0.090**	0.090**
		(0.015)	(0.016)	(0.004)	(0.043)	(0.043)
N	61,976	61,976	61,976	61,976	61,976	61,965
First stage F					84.710	84.765
Year FE	Yes	Yes	No	No	No	No
State-Year FE	No	No	Yes	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes
R^2	0.191	0.313	0.343	0.698		

Notes: This table reports the results of specifications (a) – (f) from Table 2, based on a sample excluding the top 5% of observations with the highest SSIV values. Specification (a) in this table present the first stage OLS results of the regression of incoming inventors on the IV (i.e. $\widetilde{Inv_{d,t-1,leave-out}}$). Specification (b) – (d) present OLS regressions of log (number of venture-backed startups + 1). Incoming inventors as well as the instrument are log-transformed. Specification (b) includes year fixed effects. Specification (c) includes state-year fixed effects. Specification (d) includes state-year and county fixed effects. Specifications (e) shows the results of our IV regression with state-year and county fixed effects, where incoming inventors are instrumented with $\widetilde{Inv_{d,t-1,leave-out}}$ in the first stage as defined in equation (3). Specification (f) shows results of our IV regression, but excluding the top 10 entrepreneurial counties from the sample (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). 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.

D11: Investigation of potential demand channels

To investigate potential confounds with modern day regional demand channels (e.g., certain regions are attractive to inventors with entrepreneurial ambitions), we obtained data on per capita income, the existing stock of venture-backed startups, and firm entry and exit rates. Per capita income is calculated based on yearly income level and population of counties obtained from the Bureau of Economic Analysis. The existing stock of venture-backed startups is captured by counting the venture-backed startups founded in the ten years prior to the focal year (i.e., in *t*-11 to *t*-1) in a given county, as recorded in the VentureXpert database. Firm entry and exit rates were obtained from the 2021 Business Dynamics Statistics (BDS) Datasets of the US Census.

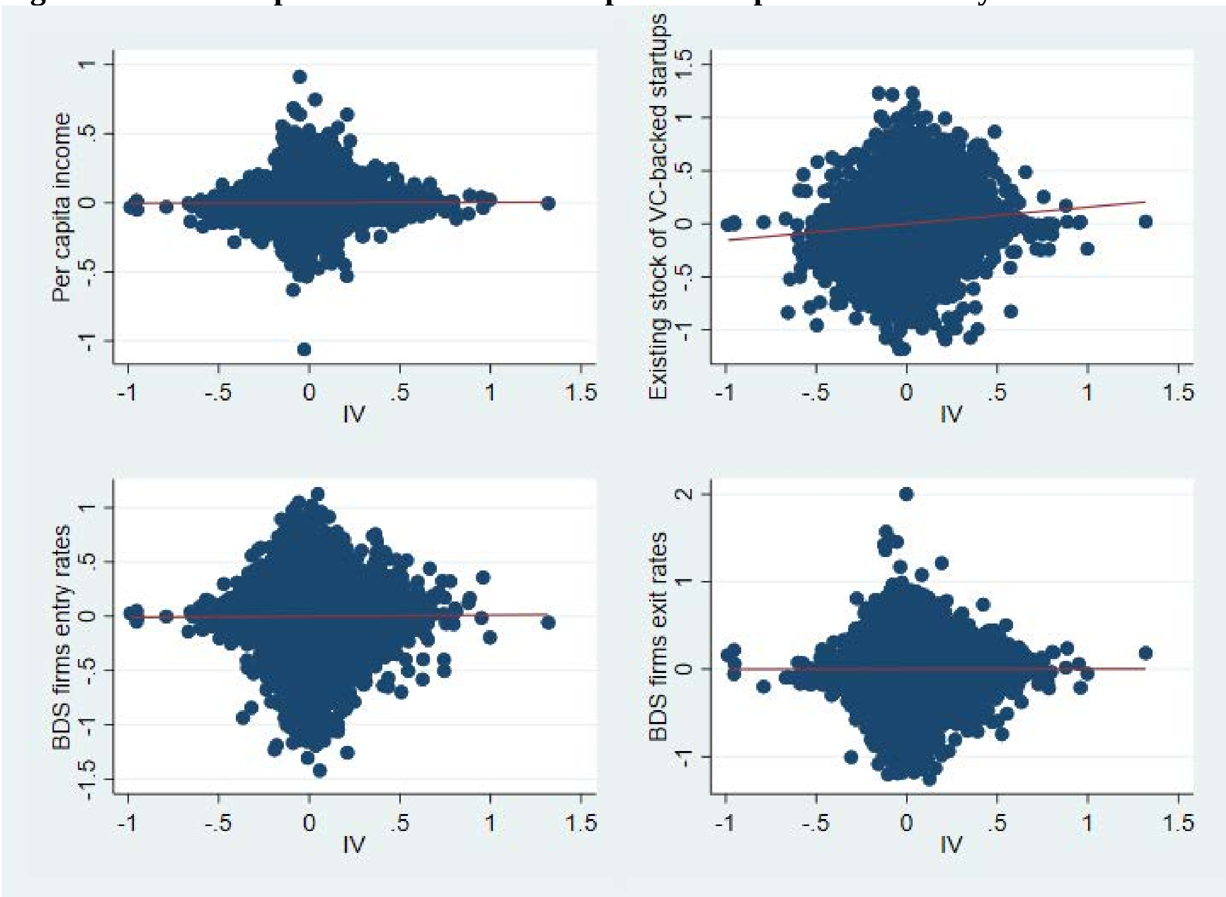
Table D11 reports regressions of the SSIV on characteristics and Figure D12 below shows corresponding scatter plots. Most relationships are insignificant, however, we find a modest but statistically significant correlation with the existing stock of startups. Besides the demand channel that would call the SSIV into question, this might occur for two reasons: a) local startup foundation rates are unlikely to randomly vary over time within county, b) our measure of local incoming inventors has a fuzzy lag structure of more than one year, i.e. a correlation with the stock may be partly driven by prior incoming inventors. The following appendix includes a control for the existing stock and re-estimates the models.

Table D11: Correlations between SSIV and entrepreneurship-relevant county characteristics

	$Inv_{d,t-1,leave-out}$
	a
	OLS
Per Capita Income$_{t-1}$	-0.007
	(0.019)
Existing stock of venture-backed startups$_{t11-1}$	0.089***
	(0.009)
Firms entry rates$_{t-1}$	0.003
	(0.002)
Firms exit rates$_{t-1}$	0.001
	(0.002)
N	62,693
State-Year FE	Yes
County FE	Yes
R^2	0.969

Notes: Per capita income is calculated based on yearly income level and population of counties obtained from the Bureau of Economic Analysis. The existing stock of venture-backed startups captures the count of all venture-backed startups founded up to 10 years prior to the focal year. BDS firm entry and exit rates were obtained from the 2021 Business Dynamics Statistics Datasets of the US Census. All variables are log-transformed and residualized by subtracting state-year and county fixed effects.

Figure D11: Scatter plot of residuals of entrepreneurship-relevant county characteristics



Notes: Per capita income is calculated based on yearly income level and population of counties obtained from the Bureau of Economic Analysis. The existing stock of venture-backed startups captures the count of all venture-backed startups founded up to 10 years prior to the focal year. BDS firm entry and exit rates were obtained from the 2021 Business Dynamics Statistics Datasets of the US Census. All variables are log-transformed and residualized by subtracting state-year and county fixed effects.

D12: Main specification re-estimated with inclusion of correlated regional characteristic

D11 uncovered a correlation between the stock of modern-day startups and the instrument. To minimize potential problems stemming from the correlation between the existing stock of venture-backed startups and the instrument for our main regressions, we re-estimated our main model controlling for the startup stock. As presented in Table D12, the results remain similar to of our main specification. However, it remains unlikely that the instrument effectively strips out all influences of modern-day demand.

Table D12: Impact of incoming inventors on local venture backed startups, controlling for the existing stock of venture-backed startups

	Incoming Inventors$_{t-1}$	Number of venture-backed startups founded				
	a	b	c	d	e	f
	OLS (first stage)	OLS	OLS	OLS	IV	IV (w/o top 10 counties)
$\widetilde{Inv_{d,t-1,leave-out}}$	0.193***					
	(0.020)					
Incoming Inventors$_{t-1}$		0.070***	0.083***	0.031***	0.202***	0.141***
		(0.008)	(0.009)	(0.004)	(0.043)	(0.038)
Existing stock of venture-backed startups$_{t-1}$	0.671***	0.322***	0.321***	0.000	-0.043***	-0.044***
	(0.015)	(0.015)	(0.015)	(0.013)	(0.012)	(0.011)
N	65,247	65,247	65,247	65,247	65,247	65,058
First stage F					169.308	159.897
Year FE	Yes	Yes	No	No	No	No
State-Year FE	No	No	Yes	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes
R^2	0.692	0.629	0.648	0.835		

Notes: This table reports the results of specifications (a) – (f) from Table 2, with controlling for the existing stock of venture-backed startups. The existing stock of venture-backed startups captures the count of all venture-backed startups founded up to 10 years prior to the focal year. Specification (a) in this table present the first stage OLS results of the regression of incoming inventors on the IV (i.e. $\widetilde{Inv_{d,t-1,leave-out}}$). Specification (b) – (d) present OLS regressions of log (number of venture-backed startups + 1). Incoming inventors as well as the instrument are log-transformed. Specification (b) includes year fixed effects. Specification (c) includes state-year fixed effects. Specification (d) includes state-year and county fixed effects. Specifications (e) shows the results of our IV regression with state-year and county fixed effects, where incoming inventors are instrumented with $\widetilde{Inv_{d,t-1,leave-out}}$ in the first stage as defined in equation (3). Specification (f) shows results of our IV regression, but excluding the top 10 entrepreneurial counties from the sample (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). 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. Note that the county-year observations with missing data on the variable controls are dropped in this analysis.

D13: Instrument remains robust to placebo shuffling by random reassignment

Given the strength of the instrument, one might be concerned that the IV absorbs unobserved local characteristics and 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). We randomly reassign computed values of the instrument for a focal county-year in three ways: (1) across the entire sample, i.e., the instrument for a focal county-year can be reassigned to any other county in any other year (2) across counties within a given year, and (3) across time within a given county (e.g., shuffling the value for Alameda County in 1994 to 2003 or 1976 but keeping it in Alameda County). Then, we re-run our baseline model with each placebo 1000 times.

Appendix Table D13 summarizes the first and second stages. All three placebos consistently show a random reassignment of instrument values effectively eliminates a significant prediction of incoming inventors in the first stage, as well as false identification of a causal impact of incoming inventors on the number of successful VC-backed startups in the second stage. Hence, our IV estimates do not appear to suffer from the artificial over-rejection of the null hypothesis as identified in other applications of shift-share instruments by Adao, Kolesár, & Morales (2019).

Table D13: Results from placebo shuffle 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.

D14: Results hold for alternative instrument constructions

We address a common inferential challenge in SSIV analyses: that particular shares may have an overly strong influence on the results. This is more of a concern with a small number of shares and may thus be ameliorated by our tracking more than 200,000 inventor surnames. Still, we investigate this by extending the temporal gap between contemporaneous inventors and those in the Census, by dropping the top 50 surnames in 1940, by entirely reformulating the instrument using the geographic centroid, and by dropping the top 5% of the distribution of modern surname frequencies. The table illustrates results consistent with the preferred instrument.

For the first alternative instrument (Table D14, model a), we consider only people in the 1940 Census who already lived in a given county before 1935. This effectively enlarges the temporal gap between the shares and the actual inventor moves, reducing potential correlation between historic and current inventor migration dynamics. In model (b), we exclude the 50 surnames that appear most frequently in 1940, addressing concerns that some families have an overly strong influence on the results. 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. Model (d) drops individuals from the 1940 Census in engineering occupations. Model (e) uses an entirely different approach, namely, the inverse distance to the geographic centroid of the name. In model (f) we aggregate the total number of mobile modern inventors and exclude the top 5% of observations in the name frequency distribution.

Table D14: Successful startups funded, using alternate instruments

	a	b	c	d	e	f
	Only individuals who settled by 1935	Drop top 50 1940s surnames	Drop wealthy families	Drop engineers	Inverse distance to the centroid	Dropped top 5% modern surnames
	IV	IV	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	0.110***	0.109***	0.106***	0.108***	0.272**	0.111***
	(0.036)	(0.035)	(0.033)	(0.033)	(0.109)	(0.032)
N	65,247	65,247	65,247	65,247	65,247	65,247
First Stage F	159.068	162.303	183.076	173.968	22.971	179.771
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	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) excludes individuals who list engineer as their occupation in the 1940 Census; (e) replaces surname shares with the inverse of distance to the centroid of each surname in 1940; (f) excludes top 5% most common surnames from modern mobile inventors. 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.

D15: Instrument remains uncorrelated with occupational change by MSA

It is possible that the instrument intended for inventors might also pick up other professions that enable entrepreneurship, e.g., managers, accountants, and lawyers. To investigate this possibility, we aggregated yearly MSA occupational data from the BLS (https://www.bls.gov/oes/tables.html, available mostly for the 2000s) and were able to calculate the change in managers, accountants, and lawyers by year and MSA (based on aggregating each occupation that had the string "Manag" or "Exec", "Account", or "Lawyer"). Given that some MSAs experienced a year to year decrease, we transformed the data using an IHS, and regressed these yearly changes on the IV. Table D15 illustrates no significant results for the non-technical professions.

Table D15: Correlation of instrument and occupational change by MSA

	a	b	c	d
	Inventors	Managers	Accountants	Lawyers
Instrument	0.439***	1.054	-0.924	-0.717
	(0.165)	(0.872)	(1.198)	(1.428)
N	1,517	1,517	1,517	1,169
Major State-Year FE	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes
First stage F	6.878	1.232	0.379	0.321
Within-R^2	0.021	0.001	0.001	0.000

Notes: This table presents OLS regressions of change in each occupation, i.e., inventor, manager, accountant, and lawyer, on the instrument. The inventor variable, i.e., Inventors, is the incoming inventor count, and all the other occupation variables, i.e., Managers, Accountants, and Lawyers, are the difference between the employment of each occupation in the current and previous years. All variables are inverse hyperbolic sine transformed. Specification (a) presents the results for inventors; (b) for managers; (c) for accountants; (d) for lawyers. Each model includes the major state-year and MSA fixed effects. Standard errors clustered at the MSA level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

D16: Comparison of OLS and IV results

As Table 2 and other results illustrate, the SSIV coefficient estimates are larger than OLS, particularly when county FE are included. We discuss three possible reasons here, including attenuation bias, measurement error, and heavier weighting of entrepreneurial hotspots by the IV. While the first two mechanisms would bias the estimates toward zero, the third cannot be dismissed, and should be recognized as a weakness of the IV.

That the OLS FE models produce much smaller coefficients might stem from attenuation bias, a common problem for FE models, e.g., because of measurement error or serial correlation. Consistent with this is the observation that the coefficients from the FE models are closer to zero irrespective of whether we estimate a positive (startups) or negative (failed startups) effect.

There is surely measurement error as well, as we infer inventor mobility from patent documents. OLS estimates will be biased towards zero in the presence of such measurement error in the main explanatory variable. Our IV estimates on the other hand are consistent even with measurement error (assuming instrument validity), which would lead to larger IV estimates compared to OLS.

One contributor to the OLS-IV gap might be that our instrument appears to be stronger for long-run high startup regions, where the elasticity between incoming inventors and startup foundations is, for various reasons, larger than in the rest of the country. This is clear because the OLS-IV gap as well as instrument strength shrinks when we exclude the top 10 counties (as measured by total amount of startup foundings over the sampling period). Hence the IV estimates for the full sample might be driven disproportionally by the hot spots of long-run startup activity.

We further investigated whether the strength of the IV varies with the ability of inventors as stronger predictive power for higher quality inventors may overweight high performers, which in turn might also explain the larger estimates from IV regressions. To shed light on this, we separated inventors by their ability, approximated by the number of future patent citations received within the previous five-year window. The first-stage F-statistic came out as 71.96 for the bottom half, whereas it was 187.94 for inventors in the upper half of the performance distribution. This shows how the IV is stronger and puts more weight on the top performing inventors. Furthermore, we observe differences in the pattern of inventor mobility based on performance. As shown in Figure D16, a county with higher startup activities tends to attract more high-performing inventors compared to low-performing inventors.

Hence the IV does not weight each county uniformly. Interpreting it as a local average treatment effect, the IV results are larger because the instrument puts higher weights on the local centers of long-run US startup activity, where more top performing inventors move to. This contributes to larger coefficients estimated with our IV estimation as compared to other models where each county receives the same weight.

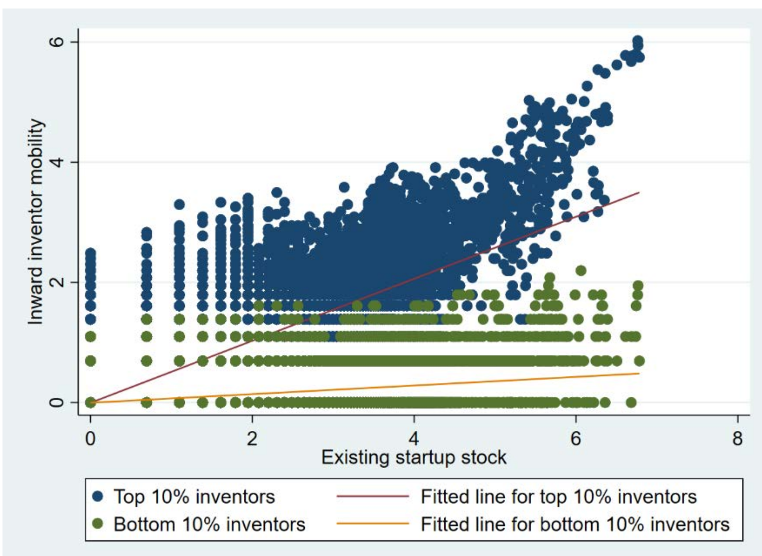


Figure D16. Inventor mobility against startup stock by performance of inventors (as measured by number of citations in five years prior to move).

Appendix E: Robustness checks and additional analyses

E1: Industry specific estimations

Table E1 provides insights into each industry (including non-tech startups). The strength of effect on the number of startups is greatest for Computers, Biotech, Communications and Media, Semiconductors, and negative for non-tech. For successful startups, it is Computer related, Communications and Media, Semiconductors, and insignificant for Biotech and non-tech. For failed startups, the effect is most negative for non-tech, also negative for Biotech and Computers, not significant for Communications and Media and Semiconductors, and positive for Computers.

Table E1: Industry specific estimations

	a	b	c
	Startups founded	Successful	Failed
Panel 1: Biotechnology + Medical/Health/Life Science			
Incoming Inventors$_{t-1}$	0.198***	00.015	-0.073***
	(0.030)	(0.013)	(0.016)
First stage F	188.672	188.672	188.672
Panel 2: Communications and Media			
Incoming Inventors$_{t-1}$	0.181***	0.050**	-0.020
	(0.046)	(0.020)	(0.016)
First stage F	119.321	119.321	119.321
Panel 3: Computer Related			
Incoming Inventors$_{t-1}$	0.390***	0.130***	0.025*
	(0.044)	(0.036)	(0.014)
First stage F	214.315	214.315	214.315
Panel 4: Semiconductors/Other Electronics			
Incoming Inventors$_{t-1}$	0.154***	0.024*	-0.008
	(0.029)	(0.013)	(0.012)
First stage F	88.459	88.459	88.459
Panel 5: Non-high-tech			
Incoming Inventors$_{t-1}$	-0.594***	-0.047	-0.624***
	(0.179)	(0.042)	(0.16)
First stage F	20.219	20.219	20.219

Notes: This table present IV regressions for each industry. Panel 1 – 5 presents the results for Biotechnology + Medical/Health/Life Science, Communications and Media, Computer Related, Semiconductors/Other Electronics, and Non-high-tech industries, respectively. Dependent variables of Columns (a)-(c) are log(number of venture-backed startups + 1), log(number of successful venture-backed startups + 1), and log(number of failed venture-backed startups + 1), respectively. 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. We define failed startups as those that are currently "Defunct" or "Bankruptcy" as indicated in VentureXpert database. 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. The number of observations is 65,247.

E2: Regressions scaled by 1940 and current populations

To increase confidence in the model specifications, Table E2 provides results scaling the left- and right-hand side variables by the 1940 and current county populations.

Table E2: Regressions scaled by 1940 and current populations

Successful venture-backed startups founded		
	a	b
	Variables scaled by 1940 population	Variables scaled by current population
	IV	IV
Incoming Inventors$_{t-1}$	0.081***	0.108**
	(0.029)	(0.043)
N	64,890	64,136
First Stage F	112.630	64.602
State-Year FE	Yes	Yes
County FE	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) and (b) present the results of regressions in which variables are scaled by log-transformed 1940 population and current population, respectively. 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. Note that the county-year observations with zero population were dropped in the weighted regressions.

E3: Regressions weighted by 1940 and current populations

To increase confidence in the model specifications, Table E3 provides results weighted by the 1940 and current county populations.

Table E3: Regressions weighted by 1940 and current populations

Successful venture-backed startups founded		
	a	b
	Weighted by 1940 population	Weighted by current population
	IV	IV
Incoming Inventors$_{t-1}$	0.116**	0.114***
	(0.039)	(0.038)
N	64,890	64,136
First Stage F	169.710	170.931
State-Year FE	Yes	Yes
County FE	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) and (b) present the results of regressions weighted by log-transformed 1940 population and current population, respectively. 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. Note that the county-year observations with zero population were dropped in the weighted regressions.

E4: Alternative model specifications: IHS transformation, time lags, inventor stocks, and growth models

Table E4 uses an Inverse Hyperbolic Sine transformation (Panel 2), 2- and 3-year lags (Panels 3-4), county controls for outward mobility, population, income, and employment (Panel 5), and growth models including inventor stocks, w/ and w/o controls (Panels 8 and 9).

Table E4: Alternative model specifications

	a	b	c
	Startups founded	Successful	Failed
Panel 1: Raw count of dependent variables			
Incoming Inventors$_{t-1}$	4.367***	0.502***	-0.216**
	(1.178)	(0.175)	(0.094)
First stage F	175.723	175.723	175.723
Panel 2: Inverse hyperbolic sine transformation			
Incoming Inventors$_{t-1}$	0.153***	0.089***	-0.194***
	(0.039)	(0.030)	(0.028)
First stage F	151.429	151.429	151.429
Panel 3: Different time lag: Two-year lag			
Incoming Inventors$_{t-2}$	0.144***	0.092***	-0.224***
	(0.036)	(0.030)	(0.026)
First stage F	204.518	204.518	204.518
Panel 4: Different time lag: Three-year lag			
Incoming Inventors$_{t-3}$	0.118***	0.082***	-0.235***
	(0.035)	(0.028)	(0.026)
First stage F	238.116	238.116	238.116
Panel 5: Main specification with controls included			
Incoming Inventors$_{t-1}$	0.420***	0.122**	-0.175***
	(0.072)	(0.050)	(0.035)
First stage F	91.094	91.094	91.094
N	54,964	54,964	54,964
Panel 6: Inventor stock			
Inventor Stock$_{t-1}$	0.143***	0.092***	-0.207***
	(0.044)	(0.033)	(0.038)
First stage F	56.668	56.668	56.668
Panel 7: Inventor stock with controls included			
Inventor Stock$_{t-1}$	0.251***	0.096*	-0.243***
	(0.080)	(0.050)	(0.063)
First stage F	23.534	23.534	23.534
N	54,964	54,964	54,964
Panel 8: Growth model			
Incoming Stock$_{t-1}$ – Incoming Stock$_{t-2}$	0.008**	0.001*	-0.002***
	(0.004)	(0.001)	(0.001)
First stage F	23.404	23.404	23.404
Panel 9: Growth model with controls included			
Incoming Stock$_{t-1}$ – Incoming Stock$_{t-2}$	0.011***	0.002*	-0.002***
	(0.004)	(0.001)	(0.001)
First stage F	23.300	23.300	23.300
N	48,854	48,854	48,854

Notes: Panel 1 uses the raw count of dependent variables. Incoming inventors and the instrument are log-transformed. Panel 2 has all DVs, incoming inventors, and instrument inverse hyperbolic sine transformed. Panels 3 & 4 present the results of the log specification with a 2- & 3-year lag applied between the DVs and incoming inventors, respectively. Panel 5 presents log-transformed results using outward mobility, population, income, and employment as controls. Panel 6 examines inventor stocks instead of inventor inflows. Panel 7 shows inventor stocks with log-transformed controls. Panel 8 has a growth model with level differences from the previous year for each DV, incoming inventors, and instrument. Panel 9 shows the growth model including log-transformed controls. DVs of Columns (a)-(c) are log(# venture-backed startups + 1), log(# successful venture-backed startups + 1), and log(# failed venture-backed startups + 1). "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. We define failed startups as those that are currently "Defunct" or "Bankruptcy" as indicated in VentureXpert database. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

E5: Different data cuts: exclusion of inventors not in 1940 Census, MSA level estimations, no entrepreneurship or mobile inventors, and winsorization

Table E5 excludes inventor names not present in the 1940 Census (Panel 1), as well as estimation at the MSA level (Panel 2), counties that never had a venture-backed startup or incoming inventor (Panels 3 and 4, respectively), and a winsorized data sample (Panel 7).

Table E5: Exclusion of inventors not in 1940 Census and sample variations

	a	b	c
	Startups founded	Successful	Failed
Panel 1: Exclude inventors whose surname does not appear in Census			
Incoming Inventors$_{t-1}$	0.185***	0.106***	-0.212**
	(0.041)	(0.033)	(0.028)
First stage F	181.869	181.869	181.869
Panel 2: Metropolitan Statistical Area (MSA) level analysis			
Incoming Inventors$_{t-1}$	0.196***	0.133***	-0.326**
	(0.056)	(0.047)	(0.053)
First stage F	82.991	82.991	82.991
N	19,047	19,047	19,047
Panel 3: Exclude counties that never had a venture-backed startup founded			
Incoming Inventors$_{t-1}$	0.356***	0.229***	-0.406***
	(0.091)	(0.076)	(0.075)
First stage F	47.612	47.612	47.612
N	21,105	21,105	21,105
Panel 4: Exclude counties that never had an incoming inventor			
Incoming Inventors$_{t-1}$	0.214***	0.127***	-0.253***
	(0.048)	(0.040)	(0.034)
First stage F	125.013	125.013	125.013
N	48,195	48,195	48,195
Panel 5: Only counties with always positive values of venture-backed startups founded			
Incoming Inventors$_{t-1}$	2.040*	2.479	-0.446
	(1.115)	(1.489)	(0.977)
First stage F	1.569	1.569	1.569
N	1,071	1,071	1,071
Panel 6: Only counties with always positive values of incoming inventors			
Incoming Inventors$_{t-1}$	1.914	2.466	-2.972
	(1.571)	(2.147)	(2.914)
First stage F	1.058	1.058	1.058
N	5,418	5,418	5,418
Panel 7: Winsorized incoming inventors			
Incoming Inventors$_{t-1}$	0.216***	0.124***	-0.254***
	(0.052)	(0.042)	(0.037)
First stage F	156.852	156.852	156.852
N	65,247	65,247	65,247

Notes: Panel 1 excludes inventors whose surname does not appear in Census data. Panel 2 presents MSA level analysis including major state-year and MSA FE. Panel 3 excludes counties that never had a venture-backed startup. Panel 4 excludes counties that never had an incoming inventor. Panel 5 includes only counties with always positive values of venture-backed startups founded. Panel 6 includes only counties with always positive values of incoming inventors. Panel 7 winsorizes incoming inventors at 99% percentile. DVs in Columns (a)-(c) are log(# VC-backed startups + 1), log(# successful VC-backed startups + 1), and log(# failed VC-backed startups + 1). "Successful" startups complete either an IPO or successful acquisition within 10 years and achieve a value ≥ 125% of total venture capital acquired. "Failed" startups are "Defunct" or "Bankruptcy" in VentureXpert. Standard errors clustered at the county level; . ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.