Innovation Spillovers Within Firms and Within Cities*

Zhao Jin† Amir Kermani‡ Timothy McQuade§ Minghao Yang¶

March 26, 2026

Abstract

We study the relative importance of within-firm and within-city agglomeration for innovation. Using novel data linking U.S. patents to detailed employment histories, we construct an inventor-level panel that tracks individuals across firms and locations. We document a positive relationship between inventor productivity and the number of inventors at the firm, alongside a negative relationship with non-inventor employment. In our most saturated specifications, a 10 percent increase in the number of inventors raises individual citation-weighted patent output by 2.2 percent. We establish causality using a mover design, a two-step AKM framework, and a shift-share instrumental variables strategy based on variation in the supply of future inventors across university programs. The IV estimates imply that 10 percent more firm inventors increase individual adjusted citations output by 1.8 percent. City-level agglomeration effects largely disappear once firm-level inventor composition is accounted for.

*We thank seminar participants at UC Berkeley.

†Cheung Kong Graduate School of Business, jinzhao@ckgsb.edu.cn

‡UC Berkeley, Haas School of Business and NBER, kermani@berkeley.edu

§UC Berkeley, Haas School of Business and NBER, tmcquade@berkeley.edu

¶UC Berkeley, Haas School of Business, mhyang@berkeley.edu

The authors of this work have no financial interests or other conflicts of interest to disclose.

1 Introduction

A central question in the economics of innovation is how the organization of economic activity shapes the production of new ideas. A large literature emphasizes the potential role of geographic agglomeration: inventors located in dense urban clusters are more productive, reflecting knowledge spillovers that arise from proximity (e.g., Jaffe, Trajtenberg, and Henderson, 1993; Moretti, 2021). At the same time, firms are the primary organizational units through which innovation is produced, raising a complementary question. That is, to what extent do spillovers operate within firms, and how do these compare to those operating across firms within cities?

This paper studies the relative importance of within-firm and within-city agglomeration forces for innovation. We combine U.S. patent records from the United States Patent and Trademark Office (USPTO) with individual-level employment histories from Revelio Labs, which compiles education and career information from online professional profiles. Linking these two datasets allows us to track inventors over time, across firms, and across locations. The merged data yield an inventor-level panel covering nearly 500,000 inventors over four decades, with detailed information on coworkers, firm composition, and geographic location. This structure allows us to disentangle the effects of firm-level and city-level agglomeration on individual innovative output in a unified empirical framework.

We begin by revisiting a well-established empirical regularity: inventor productivity declines with total firm size. Consistent with prior work, we find that both patents per employee and patents per inventor decline as total firm employment increases. However, total employment conflates two distinct forces—exposure to other inventors, which may generate knowledge spillovers, and exposure to non-inventor employees, which may reflect organizational complexity and bureaucratic frictions. Separating these forces reveals a new fact: inventor productivity increases with the number of other inventors in the firm, but decreases with the number of non-inventors.

In our most saturated specification with inventor, firm, and city-field-year fixed effects, a 10 percent increase in the number of inventors at a firm increases an individual inventor's citation-weighted patent output by approximately 2.2 percent, while a comparable increase in non-inventor employment reduces output by about 0.7 percent. Decomposing inventors by research field shows that these effects are highly localized. The positive spillover is driven almost entirely by same-field inventors, with little to no effect from inventors working in other technological areas. These results suggest that the negative relationship between firm size and innovation masks a positive

within-firm agglomeration force among inventors that is offset by the organizational costs of scale.

Of course, a central challenge in interpreting these estimates is distinguishing causal spillovers from sorting and other confounding factors. We address these concerns using three complementary empirical strategies. First, we exploit within-inventor variation using a mover design. When inventors move between firms, we observe sharp and persistent changes in their innovation output that track differences in the destination firm's environment. Inventors who move to firms with higher inventor density experience immediate increases in patenting that persist for several years, while those moving to less dense environments experience declines. Importantly, we find no evidence of differential pre-trends, suggesting that these changes are not driven by anticipatory behavior or differential trajectories prior to the move.

Second, we implement a two-step Abowd-Kramarz-Margolis (AKM) framework that separates inventor fixed effects from firm-year productivity components. This approach allows firm-level innovation environments to evolve over time and uses mobility across firms to identify these effects. In the second stage, we show that firm-year productivity is increasing in the number of inventors at the firm, even after controlling for firm and year fixed effects. The magnitudes are consistent with the baseline estimates, reinforcing the interpretation that inventor density within the firm causally affects productivity.

Third, we develop an instrumental variable (IV) strategy that exploits plausibly exogenous variation in the supply of future inventors across university programs. The key idea is to construct a shift-share instrument for a firm's inventor share based on (i) its historical hiring patterns across university-degree-major groups and (ii) time-varying shocks to the propensity of graduates from those programs to become inventors. Specifically, we fix each firm's exposure to different educational pipelines using its 2009 workforce composition, and then interact these shares with national changes in inventor supply across programs over time. To address concerns about endogenous matching between firms and educational programs, we implement a leave-one-out procedure that excludes each firm's own hires when constructing the supply shocks.

This instrument generates strong first-stage variation (F-statistic above 900) and yields causal estimates that are somewhat smaller than the OLS estimates but remain economically meaningful. In the IV specification, a 10 percent increase in the number of inventors at a firm increases individual innovation output by approximately 1.8 percent. The fact that the IV estimates are smaller than the OLS estimates suggests that endogenous hiring responses to positive productivity shocks may bias the OLS upward, but the persistence of a sizable effect confirms the presence of causal

within-firm spillovers.

Having established the importance of within-firm agglomeration, we turn to the role of geographic agglomeration. Consistent with prior work, we initially document that inventors located in cities with more same-field inventors are more productive. In specifications without firm controls, the semi-elasticity of inventor productivity with respect to city-level same-field inventor density is approximately 0.08 to 0.11. Once we add inventor and firm fixed effects, we confirm that the city-level effect is robust, indicating that the effect is not driven by the sorting of more productive inventors or firms into larger cities.

Next, to directly compare firm-level and city-level spillovers, we estimate horse-race specifications that include both sets of variables simultaneously. In these regressions, the semi-elasticity of inventor productivity with respect to within-firm same-field inventor density remains stable at approximately 0.06 and highly statistically significant, while the corresponding semi-elasticity for city-level inventor density now falls to near zero. This pattern suggests that the apparent productivity benefits of increases in the number of inventors within a city are, in fact, due to higher within-firm innovation spillovers arising from increases in the number of inventors working at the firms in that city.

We provide additional evidence on the mechanisms underlying these results. First, we show that within-firm spillovers are particularly strong among active and high-productivity inventors, consistent with knowledge exchange rather than pure scale effects. Second, we decompose forward citations by their source and find that within-firm citations account for a disproportionately large share relative to the size of the potential citing pool, indicating that knowledge flows are concentrated within organizational boundaries. Third, we show that inventors exposed to more colleagues are more likely to found startups, suggesting that within-firm interactions generate economically meaningful knowledge that extends beyond the firm.

Taken together, our findings point to the firm as the primary locus of innovation spillovers. While cities facilitate the matching of inventors and firms and may shape the equilibrium distribution of innovative activity, the productivity gains from agglomeration appear to operate predominantly through interactions within firm boundaries. An important implication is that estimates of geographic spillovers that do not account for firm structure may overstate the role of cities by conflating spatial clustering with organizational composition.

1.1 Literature Review

Our paper contributes to several strands of literature. First, we relate to the research on innovation and firm size. Acs and Audretsch (1988) document that small firms, measured by employment, often generate more innovations per employee and per dollar of R&D than large firms, especially in innovative industries. Akcigit and Kerr (2018) and Akcigit and Goldschlag (2025) confirm this negative relationship between total firm employment and individual inventor patenting output. A large body of literature has also examined various mechanisms that might cause large, established firms to exhibit weaker incentives for innovation, especially radical innovation, including, for example, market power (e.g., Aghion et al., 2005), cannibalization effect (e.g., Arrow, 1962; Igami, 2017), and agency problem (e.g., Edmans, Fang, and Lewellen, 2017; Seru, 2014). While we confirm this negative relationship regarding total employment, we document a novel positive relationship between inventor productivity and the total number of inventors at the firm.

Second, we contribute to the literature on geographic agglomeration. Prior studies suggest that innovation could benefit from knowledge spillovers arising from the concentration of economic activity.[1] Jaffe, Trajtenberg, and Henderson (1993) compared the locations of cited and citing patents to show that citations are far more likely to come from the same metro area (SMSA) or state than would be expected by chance. Moretti (2021) provides micro-level evidence showing that when an inventor relocates to a city with a larger concentration of same-field inventors, their number of patents and patent citations rises significantly. This implies that inventor agglomeration in cities translates into individual innovation productivity gains. Our contribution to this strand is showing that, once firm-level spillovers are taken into account, these city-level effects largely reflect the sorting of more productive inventors and firms into technology clusters.

Third, we build on the study of within-firm knowledge spillovers. Henderson and Cockburn (1996) provide direct evidence of substantial within-firm innovation spillovers. Focusing on the pharmaceutical industry, they find that large firms enjoyed higher research productivity because they could internally transfer knowledge between different project teams. Furthermore, Giroud, Liu, and Mueller (2024) show that when a firm has R&D sites in multiple cities, a big innovation cluster like Silicon Valley not only boosts its local inventors, but also increases the productivity of the firm's inventors in other, distant cities through internal networks. We extend this research by utilizing a mover design to provide evidence of positive agglomeration forces operating specifically

[1] See, for example, Marshall (1890), Griliches (1958), Lucas (1988), Jovanovic and Rob (1989), Saxenian (1996), Glaeser (1999), Moretti (2004a), Moretti (2004b), and Greenstone, Hornbeck, and Moretti (2010).

within the firm.

Finally, we relate to the research on social networks and knowledge diffusion. Singh (2005) shows that knowledge spillovers are strongly shaped by co-inventorship networks. Inventors who have previously collaborated are significantly more likely to cite each other's patents later. Azoulay, Graff Zivin, and Wang (2010) examine the unexpected deaths of star scientists and find that their collaborators' future output drops significantly, highlighting how the loss of a key local knowledge source negatively impacts peers' innovation. Almeida and Kogut (1999) show that knowledge tends to flow along paths of labor mobility and document that an inventor was more likely to cite a patent of a former employer or a firm that had hired talent from his current company. Atkin, Chen, and Popov (2022) exploits smartphone geolocation data to directly observe encounters between inventors of different firms within Silicon Valley. They find that when two innovators meet in person, their companies subsequently exchange significantly more patent citations. Our paper contributes to this strand by investigating the relative magnitude of these spillovers, finding that the firm boundary is a primary driver of the innovation gains often attributed to broader urban networks.

2 Data and Sample Construction

This section discusses our data sources, the process we use to construct our sample, and presents summary statistics at the inventor, the firm, and the city levels. The data used in the analysis mainly come from two sources. The first is US patent data from the United States Patent and Trademark Office (USPTO). The second is detailed employment history data from Revelio Labs, which include public education and employment information disclosed by workers. Merging these datasets allows us to track innovators' full employment records, including when and where they move and how many colleagues they work alongside.

2.1 Patent Data

We extract all US utility patents granted between 1976 and 2022 from PatentsView.org[2]. For each patent, we obtain information on the application date, grant date, inventor names and addresses,

[2]See https://patentsview.org/download/data-download-tables

assignee names and addresses, IPC and CPC classes[3], forward citations, and full text.

The raw patent data do not contain unique inventor, assignee, or location identifiers, so it is not possible to accurately track inventors over time. PatentsView.org conducts a disambiguation process that clusters similar inventor, assignee, and location records to assign persistent IDs.[4] This allows us to track inventors over time, conditional on observing them on at least one patent in a given year.

We restrict our sample to patents filed by at least one US inventor, leaving us with roughly half of all grants. We assign each inventor to a city based on the county of residence that appears most frequently across her patents in that year; following Moretti (2021), we define a city as a Bureau of Economic Analysis (BEA) "economic area." We also assign each inventor to one of five research fields based on her modal IPC class; we define research fields as in Moretti (2021).[5]

2.2 Employment History Data

We obtain individual-level employment histories from Revelio Labs. As of early 2023, the dataset contains more than 450 million individuals from 180 countries. Individuals self-report their education and employment information, including (but not limited to) the institutions they attended, degrees obtained, cities of residence, employers, and job titles. For each experience, they may report start and end dates.

Employment histories are available only if an individual chooses to post them, so coverage varies substantially by occupation, industry, and time. Individuals with college education and those in white-collar jobs are more likely to publish employment histories. Because this paper focuses on US inventors—most of whom have college education and work in white-collar jobs—selection is likely to be less of a concern.

[3]The International Patent Classification (IPC), maintained by the World Intellectual Property Organization (WIPO), and the Cooperative Patent Classification (CPC), developed jointly by the USPTO and the European Patent Office (EPO), label patents by their technology area, with CPC typically providing finer detail.

[4]See https://patentsview.org/disambiguation; and https://patentsview.org/disambiguation/disambiguation-metrics

[5]Darby and Zucker (1999) and Zucker and Darby (1999) provide a crosswalk between IPC classes and research fields.

2.3 Dataset Merge

We aim to construct an inventor-year panel that tracks inventors from the start of their careers through retirement. This panel enables us to identify the exact year in which inventors move, as well as the location and firm associated with each move. To build the panel, we merge the patent data with employment history data, which allows us to observe inventors beyond their patenting years.

The merge procedure uses inventors' first and last names, the assignee names or CUSIPs[6], assignee locations, and firms' parent-subsidiary relationships. We require that the inventor name match the name in the employment history data, conditional on the assignee matches the employer reported by the worker. We also observe that the patent assignee and the worker's reported employer are often a parent-subsidiary pair, and we treat such cases as matches to increase the match rate. We delegate the detailed matching procedure to the Appendix. This procedure matches approximately 500,000 inventors, accounting for about one-third of all US inventors from 1979 to 2022. The matched sample is not a random draw from the population of US inventors: the match rate is higher for inventors at larger firms (44% for firms with 1,000+ employees versus 42% for firms with 40+ employees), for more prolific inventors (54% for those with 10+ patents versus 40% for those with 2+), and for more recent patents (42% post-2010 versus 17% pre-2000). The time pattern reflects the growth of online professional profiles, while the productivity pattern reflects the greater visibility of prolific inventors. Table IA1 in the Internet Appendix reports detailed conditional match rates, and Figures IA1–IA3 provide further detail across subgroups, including by inventor patent count (Figure IA2). While we cannot fully rule out that the nonrandom match rate biases our estimates, the robustness of results across different subsamples and specifications is reassuring.

In the end, the merged dataset provides annual employment records of inventors. We first determine the firm the inventor works for, and the city the inventor lives in. The inventor's firm is defined as that reported by the inventor in the employment history. We assign the inventor's city using the following order of precedence: the inventor's patent record, the inventor's reported place of residence, and (if still missing) the firm's location. If an inventor changes employers or relocates, we define her firm and city as the most updated ones. We assign an inventor's research field in a given year as the modal field across all patents she applies for in that year. If an inventor

[6]CUSIPs are available only for public firms. For patent assignees, CUSIPs come from Kogan et al. (2017) and Compustat. For employment history data, Revelio Labs provides a CUSIP for each public firm.

does not apply for any patents in that year, we carry forward her most recent field; if she has not yet applied for any patents, we assign her the field from her first patent(s).

We then add patent counts and citations to the panel. To adjust for different innovation and citation patterns over time and across classes, we follow Lerner and Seru (2022) and measure normalized citations as the number of citations a patent received scaled by the average number of citations for other patents in the same year and CPC class. To avoid outliers, we winsorize patent counts and normalized citations at 99.9 percentile. We use normalized citations as the primary outcome of interest.

Next, we compute the number of inventors and non-inventors in each firm-year. Within a firm, we classify an employee as an inventor if she has filed at least one patent at any point in time. For each inventor, we further split the inventor headcount in her firm into inventors in her research field and inventors outside her field. We construct analogous inventor measures at the city-year level.

Part of our identification strategy relies on a mover design. We define an inventor as a "non-mover" if she continues to work at the same firms in a year. We define an inventor as a "mover" if she switches to another firm in that year.

2.4 Summary Statistics

Panel A of Table 1 presents summary statistics at the inventor-year level. The sample covers matched inventors from 1979 to 2019. For each inventor-year, we report the key outcome variable adjusted citations. We also report the patent counts as well as working experience. Panel B presents summary statistics at the firm-year level, including the total number of employees, the number of inventors, the number of non-inventors, and average adjusted citations per inventor. Panel C presents summary statistics at the city-field-year level, including the number of same-field inventors and other-field inventors.

3 Within-Firm Innovation Spillovers

In this section, we examine the relationship between inventor productivity and the composition of a firm's workforce. We begin by documenting binscatter evidence, then present OLS regression results, and finally employ a two-step AKM estimation and mover design to address endogeneity concerns.

3.1 Baseline estimates

A large literature has documented a negative relationship between total firm employment and patents per employee (e.g., Acs and Audretsch, 1988; Akcigit and Kerr, 2018; Akcigit and Goldschlag, 2025). Figure 1 replicates this finding in our data: both patents per employee (left panel) and patents per inventor (right panel) decline with total firm employment. However, total employment conflates two groups that may have opposite effects on inventor productivity: fellow inventors, who may generate knowledge spillovers, and non-inventors, whose presence proxies for organizational scale and bureaucracy.

Figure 2 separates these two effects. The left panel shows that, conditional on the number of inventors, patents per inventor decline with the number of non-inventors. The right panel shows that, conditional on the number of non-inventors, patents per inventor *increase* with the number of inventors.[7] This suggests that the negative relationship between firm size and inventor productivity masks a positive within-firm spillover among inventors.

3.1.1 The OLS estimates

To formalize these relationships, we estimate the following OLS specification:

$$y_{i,f,c,k,t} = \beta_1 \ln(Firm_Inv_{f,t}) + \beta_2 \ln(Firm_NonInv_{f,t}) + \alpha_{c,k,t} + \phi_i + \mu_f + \varepsilon_{i,f,c,k,t}, \qquad (1)$$

where $y_{i,f,c,k,t}$ is the number of citation-adjusted patents of inventor i working at firm f in city c and research field k in year t. The variable $Firm_Inv_{f,t}$ is the number of inventors at firm f in year t, and $Firm_NonInv_{f,t}$ is the number of non-inventors. The term $\alpha_{c,k,t}$ denotes city-field-year fixed effects, ϕ_i denotes inventor fixed effects, and μ_f denotes firm fixed effects. Equation (1) shows the most saturated specification; we progressively build toward it, starting with city-field and field-year fixed effects and sequentially adding inventor, city-field-year, and firm fixed effects. Standard errors are clustered at the city-field by field-year level.

Table 2 presents the results. Across all specifications, β_1 is positive and statistically significant, indicating that inventors are more productive when they work alongside more fellow inventors. In contrast, β_2 is negative and significant, confirming that a larger non-inventor workforce is asso-

[7] Figure IA5 in the Internet Appendix further shows that the positive (negative) relationship between patents per inventor and the number of inventors (non-inventors) is visible even without controlling for the other group, and Figure IA6 documents the positive correlation between the number of inventors and non-inventors within firms.

ciated with lower inventor productivity. The estimates are stable as we progressively saturate the model with fixed effects. In the most demanding specification with inventor, city-field-year, and firm fixed effects, a 10% increase in the number of inventors is associated with a 2.2% increase in adjusted citations relative to the mean, while an increase in non-inventors is associated with a decrease of 0.7% relative to the mean.

3.1.2 Robustness Checks

We confirm that our baseline results are robust to a range of alternative specifications and sample selections. Table IA2 in the Internet Appendix shows that the results hold when using raw patent counts instead of adjusted citations as the dependent variable. Table IA3 reports results using ln(adjusted citations). Because citation-adjusted patents equal zero in inventor-years without any patent applications, our dependent variable captures both the extensive margin and the intensive margin. We examine these margins separately. Table IA5 restricts the sample to the intensive margin—inventor-years with at least one patent—and finds similar results, confirming that within-firm spillovers operate on both the extensive margin (whether to patent) and the intensive margin (citation quality conditional on patenting).

Table IA6 uses only patents that are matched to employment histories. Table IA7 decomposes the inventor count into active inventors (those who patented in the past five years) and idle inventors, and shows that the spillover effect is concentrated among active inventors. Across all these specifications, the positive effect of the number of inventors and the negative effect of non-inventors remain qualitatively and quantitatively similar. Figure IA7 in the Internet Appendix further replicates the binscatter evidence using inventor counts from PatentsView data instead of matched employment histories, confirming that the patterns are not an artifact of the matching procedure.

3.1.3 Inventor field

If the positive association between inventor headcount and productivity reflects knowledge spillovers, the effect should be strongest among inventors working on related technologies, who are most likely to exchange relevant ideas. To examine this, we further decompose the inventor count into same-field inventors and other-field inventors. Table 3 shows that the positive spillover is driven primarily by inventors in the same research field. In the most saturated specification, the coefficient

on ln(*Firm_FieldInv*) is 0.060 ($p < 0.01$), while the coefficient on ln(*Firm_OtherInv*) is -0.001 and statistically insignificant. The effect of non-inventors also becomes insignificant once field-specific inventor counts are included. This is consistent with knowledge spillovers being more effective among inventors working on related technologies. Table IA10 in the Internet Appendix examines additional dimensions of heterogeneity.

3.1.4 Inventors: stayer vs. mover

A natural concern with the OLS estimates is that they reflect endogenous sorting: more productive inventors may selectively move to firms with more fellow inventors. If sorting were the primary driver, the positive coefficient should be larger for movers—who choose their destination firms—than for stayers, whose variation in ln(*Firm_FieldInv*) comes only from within-firm changes over time. Table 4 tests this by estimating the specification separately for each group. For stayers, we look at inventors who stayed at the same firm for at least three years, until they moved to another firm. For movers, we look at inventors who moved in a given year, and include a three-year window around the moving year. The coefficient on ln(*Firm_FieldInv*) is nearly identical for stayers (0.056) and movers (0.057), both significant at the 1% level, suggesting that the positive association is not primarily driven by selection in the moving decision.

3.1.5 Additional analyses for knowledge spillovers

Four additional analyses support the interpretation that the within-firm association reflects knowledge spillovers rather than a pure scale effect. First, Table IA7 in the Internet Appendix decomposes the inventor count into active inventors and idle inventors, where active inventors are those who applied for at least one patent in the past five years, and idle inventors are those who did not. In the most saturated specification, the coefficient on ln(*Firm_ActiveInv*) is 0.055 ($p < 0.01$), while the coefficient on ln(*Firm_IdleInv*) is smaller (0.026). This is consistent with knowledge spillovers requiring ongoing research engagement: inventors who are actively producing knowledge are more useful peers than those who have stopped patenting. Second, we are able to explicitly control for firm age, research and development (R&D) expenditure, as well as selling, general, and admin (SG&A) expenditure for public firms. Table IA9 in the Internet Appendix shows the result. When no additional controls are added, the results in column (1) are comparable to that in column (4) of Table 3. Meanwhile, controlling for firm age square, R&D expenditure, or SG&A expenditure do

not meaningfully change the coefficient estimate on the log number of field inventors at the firm. This indicates that the positive relationship is not driven by firm life-cycle differences or higher research investment, but may reflect knowledge spillovers from other inventors within the firm. Third, we use employees' education records and job information to predict whether they are inventors[8], and then construct firm-level counts of predicted inventors and non-inventors. Table IA8 in the Internet Appendix reports the results. Comparing columns (2)–(3), which use predicted inventor and non-inventor counts, with column (1), which uses actual counts, we find that the estimate of β_1 is smaller (approximately 0.035 versus 0.077) but remains positive and statistically significant. This suggests that the knowledge spillover effect persists even when we adopt an alternative, prediction-based definition of who engages in innovation. Fourth, Table IA10 in the Internet Appendix looks at spillover effects among star inventors (inventors above the ninetieth percentile in the total number of patents). It shows that within-firm spillovers are larger at large firms (coefficient of 0.107, relative to 0.076 in the full sample) , consistent with knowledge exchange being more valuable in environments with richer technological opportunities. Table IA10 also shows that within-firm spillovers are stronger among star inventors (0.157, relative to 0.076 in the full sample). Star inventors are the most research-intensive employees, and their productivity is most sensitive to the research environment around them, which is consistent with knowledge exchange rather than a pure scale effect driving the results.

We note several identification challenges inherent in interpreting these OLS estimates as causal spillovers. First, changes in inventor headcount may alter the composition and average ability of a firm's inventor pool, confounding spillovers with selection. Second, the estimates may reflect reverse causality: firms that experience positive productivity shocks may simultaneously hire more inventors and produce more patents. The inventor and firm fixed effects in equation (1) address time-invariant heterogeneity, but they do not fully resolve time-varying confounders. We address these concerns more directly with a two-step AKM design and an instrumental variables design.

3.2 Two-step AKM and mover design

We start with an extension of the Abowd-Kramarz-Margolis (AKM) framework (Abowd, Kramarz, and Margolis, 1999). This framework allows us to separate persistent inventor ability from firm-

[8] We estimate logistic regressions of the inventor indicator on employees' university, major, degree, job seniority, job title, and job description, and classify those with the highest predicted probabilities as inventors. See the Internet Appendix for details.

specific productivity components while accounting for sorting between inventors and firms. A key restriction in the canonical AKM framework is that firm effects are time-invariant. In the context of innovation spillovers, however, this restriction is likely to be violated. Firms experience substantial changes in research focus and employment over their life cycle, which can affect both inventor productivity and the number of inventors at the firm. If such changes are not accounted for, a time-invariant firm effect may confound persistent firm differences with the life-cycle evolutions of the firms, potentially biasing the estimated innovation spillovers.

To accommodate these features, we extend the AKM framework by allowing firm effects to vary over time. This approach follows Lachowska et al. (2023) that relaxes the time-invariance restriction in order to capture evolving firm conditions. As in the canonical AKM framework, identification relies on movers to separate individual and firm components and to distinguish firm effects across different firms. Stayers, in contrast, are used to identify the time variation in each firm component through within-firm changes over time. Allowing firm effects to vary over time enables us to capture the evolution of inventor productivity across firms and relate those changes to within-firm shifts in the number of inventors.

In the first step, we estimate

$$y_{i,f,c,k,t} = \phi_i + \eta_{f(i,t),t} + \alpha_{c(i,t),k(i,t),t} + x'_{i,t}\gamma + \varepsilon_{i,f,c,k,t}, \tag{2}$$

where ϕ_i represents the inventor fixed effects, $\eta_{f,t}$ represents the firm-year fixed effects, and $\alpha_{c,k,t}$ denotes the city-field-year fixed effects. This specification isolates firm-year level productivity components ($\eta_{f,t}$) net of inventor heterogeneity (ϕ_i) and city-level spillovers ($\alpha_{c,k,t}$).

In the second step, we regress the estimated firm-year effects on the size of the firms:

$$\eta_{f,t} = \beta_1 \ln(Firm_Inv_{ft}) + \beta_2 \ln(Firm_NonInv_{ft}) + \mu_f + \tau_t + u_{f,t}. \tag{3}$$

where μ_f represents the firm fixed effects and τ_t represents the year fixed effects.

Table 5 displays the results from the second-step estimation. Column (1) does not control for firm or year fixed effects; Column (2) controls for year fixed effects; Column (3) controls for firm and year fixed effects. The standard errors are clustered at firm level. The coefficient estimates for the number of inventors are consistently positive and comparable across specifications, confirming that inventor productivity increases with the number of other inventors in the firm.

A key identification assumption for the AKM framework is that the timing of moves is random.

To test this assumption, we follow the event study design of Finkelstein, Gentzkow, and Williams (2016) and examine the productivity changes around moves of inventors.

For a moving inventor i whose origin and destination firms are $o(i)$ and $d(i)$, respectively, we define by δ_i the difference in average innovation output between the inventor's destination and origin:

$$\delta_i = \bar{y}_{d(i)} - \bar{y}_{o(i)}, \tag{4}$$

where $\bar{y}_{d(i)}$ is computed by averaging over all inventor-years (including non-moving inventors) at the destination firm (and analogously for $\bar{y}_{o(i)}$ at the origin firm). This measure captures the difference in firm components implied by the move. The left panel of Figure 3 shows the distribution of these move sizes for all movers in our sample.

We then estimate an event study equation:

$$y_{i,t} = \alpha_i + \theta_{r(i,t)}\delta_i + \tau_t + x'_{i,t}\gamma + \varepsilon_{i,t}, \tag{5}$$

where ϕ_i represents the individual fixed effects, $\theta_{r(i,t)}$ are event-times relative to inventor i's move, τ_t represents the year fixed effects, and x_{it} represents a set of control variables.

The right panel of Figure 3 shows the event study results. We plot the event-time coefficients ($\theta_{r(i,t)}$) from Equation (5). Since these coefficients are only identified up to a constant term, we normalize $r(i,t) = -1$ to be zero. The results suggest that prior to the move, there are no clear trends in inventor productivity, so the innovation spillovers are unlikely to reflect anticipatory behaviors or differential productivity dynamics. Moreover, there is a sharp and persistent jump in inventor productivity in the move year and the two years afterward. Inventors who move to firms with higher patent productivity experience an immediate increase in patent output. After the initial climb, the coefficients remain fairly stable in the 0.75 to 1 range.

Figure 4 presents two event studies separately for inventors moving to a higher-productivity firm (left panel) and inventors moving to a lower-productivity firm (right panel). We observe positive coefficients in both panels, consistent with movers increasing or decreasing their patent productivity depending on the direction of the move. Moreover, the coefficient magnitudes are comparable, indicating that the role of inventor-firm sorting may be limited.

The previous event studies verify that there are no pre-trend before the move, and the inventor productivity increases when the inventor moves to a higher-productivity firm. A natural question would be: How much of the effects are driven by innovation spillovers? To investigate this, we

conduct an alternative event study:

$$y_{i,t} = \phi_i + \theta_{r(i,t)}\delta_{size,i} + \tau_t + x'_{i,t}\gamma + \varepsilon_{i,t}, \tag{6}$$

where we define $\delta_{size,i}$ by the log difference in average number of inventors between the inventor's destination and origin at the time the inventor worked there:

$$\delta_{size,i} = \ln(\overline{Firm_Inv_{d(i)}}) - \ln(\overline{Firm_Inv_{o(i)}}). \tag{7}$$

The coefficient estimates illustrate the dynamic responses of inventor productivity when the inventor's move causes a change in the number of other inventors at the firm. The left panel of Figure 5 presents the distribution of $\delta_{size,i}$.

The right panel of Figure 5 presents the event study result. There exists a small pre-trend before the inventor's move, but the trend is never significant. Meanwhile, when an inventor moves to a firm with more inventors, her inventor productivity hikes for the first four years after the move, and remains relatively stable. This pattern is consistent with our previous regression results in Table 3, and further suggests that the spillovers are persistent.

The event-study results are robust to several alternative approaches. First, following Card, Rothstein, and Yi (2025), we divide movers based on average productivity quartiles of their origin and destination firms, and plot the average inventor productivity for moves across different quartiles. Figure IA8 in the Internet Appendix presents the results. The raw data display clear discrete changes at the time of the move and exhibit parallel trends prior to the move, confirming that the main event study results.

Second, we replicate the event studies using a subset of patents associated with new technologies defined as in Kalyani et al. (2025). Figures IA9–IA11 in the Internet Appendix present the results. The patterns are similar, if not larger. This finding is consistent with the view that spillovers within firms are particularly important for innovation at the technological frontier.

3.3 Instrumental Variables Design

Another concern is that time-varying innovation shocks at the firm level may simultaneously increase inventor headcount and per-inventor patent output. In this case, OLS estimates would be upward biased. To address this concern, we construct a shift-share instrument for the number of

inventors within firms. The instrument exploits plausibly exogenous variation in the labor supply shocks of future inventors across university programs. We define the "shift" as the probability that a graduate from a given university program becomes an inventor in the future, and the "share" as each firm's hiring distribution across these university programs. To avoid simultaneity with subsequent inventor hiring, we construct shares using firm employment in 2009, and hold those shares fixed while allowing shifts to vary from 2010 to 2019. We multiply each group's shift by the firm's baseline hiring share and aggregate across all feeder university-major-degree groups to obtain a firm-level measure of the predicted probability of inventors.

To construct the instrument, we rewrite the number of inventors at firm f in year t as $Firm_Inv_{f,t} \equiv Firm_Emp_{f,t} \times Pr_Inv_{f,t}$, where $Firm_Emp_{f,t}$ is the number of employees at firm f in year t, and $Pr_Inv_{f,t}$ is the probability that an employee at firm f is an inventor in year t. We instrument $Pr_Inv_{f,t}$ as

$$\widehat{Pr_Inv}_{f,t} = \sum_{p \in \mathscr{P}} w_{p \to f}^{2009} \times \frac{\sum_{f' \neq f} Grad_Inv_{p \to f',t}}{\sum_{f' \neq f} Grad_Emp_{p \to f',t}}. \tag{8}$$

Here p denotes a university program, which is a university-degree-major combination that the employees report[9]. $w_{p \to f}^{2009}$ is the share of employees at firm f in 2009 who graduated from program p, $Grad_Inv_{p \to f',t}$ is the number of future inventors graduating from program p and then working in firm f' in year t, and $Grad_Emp_{p \to f',t}$ is the number of employees graduating from program p and then working in firm f' in year t. We then define the predicted number of inventors as

$$\widehat{Firm_Inv}_{f,t} = Firm_Emp_{f,t} \times \widehat{Pr_Inv}_{f,t}, \tag{9}$$

and re-estimate

$$y_{i,f,c,k,t} = \beta_1 \ln(Firm_Inv_{f,t}) + \alpha_{c,k,t} + \phi_i + \mu_f + \varepsilon_{i,f,c,k,t}, \tag{10}$$

with the instrumented number of inventors. When we estimate this equation, we always work on the most saturated version because the firm fixed effects help alleviate the impact of firm's employment composition on inventor productivity[10].

[9] To ease computation, we group degrees into PhD degrees versus non-PhD degrees. and majors into Science, Technology, Engineering, and Mathematics (STEM) majors versus non-STEM majors.

[10] Table IA11 in the Internet Appendix compares the OLS estimates from equation (1) and (10), where we find that β_1 from equation (1) is comparable to that from equation (10) when firm fixed effects are controlled. This suggests that including firm fixed effects in equation (10) helps alleviate the effect of firm employment composition that would otherwise bias the estimates.

The exclusion restriction of our instrument relies on the conditional orthogonality of inventor supply shocks from university programs to inventor productivity confounders at firms (Borusyak, Hull, and Jaravel, 2022). Specifically, in a given year, a university program where more future inventors graduates proportionally is not systematically connected to firms which, at the same time, experiences higher inventor productivity for reasons other than spillover effects through fellow inventors. A threat to this assumption is the sorting between university programs and firms, and we adopt two strategies to accommodate this concern. First, when we construct the "shift" for a given firm f, we exclude the graduates that are working in that firm. This leave-one-out method breaks the direct link between programs and firms, and relies on the variation from programs. Second, including the firm fixed effects in equation (10) controls for firm specific invention environment and/or productivity shocks that are attracting or driving away potential inventors.

Table 6 shows the instrument variable regression results. Column (1) shows the OLS result restricted to firms for which the instrument can be constructed and to years 2010–2019. Column (2) shows the IV result. Column (3) shows the first-stage result of the IV regression. In all specifications, we control for city-field-year fixed effects, inventor fixed effects, and firm fixed effects. The standard errors are clustered at city-field by field-year level. The OLS coefficient in Column (1) is larger than that in Table 2 because the sample period is restricted to 2010 to 2019, when the average per-inventor patenting becomes larger than earlier years (0.418 vs. 0.349). The IV coefficient in Column (2) is 0.075 and is significant at the 1% level. The first-stage F (963) is well above conventional weak-instrument thresholds, alleviating the weak instrument concern. In terms of magnitude, a 10% increase in the number of inventors is associated with a 1.8% increase in adjusted citations relative to the mean. Notably, the IV coefficient is smaller than the OLS coefficient (0.075 vs. 0.109), indicating that the OLS estimates may be biased upwards. Column (3) confirms a strong first stage: the coefficient on predicted log number of inventors is 0.792 (t-statistics > 30), and the regression has a R^2 above 0.99[11]. Overall, the IV results support a causal interpretation of spillover effects.

Our IV results are robust to alternative specifications. First, Table IA13 in the Internet Appendix shows similar results when we define the shares from 2005 and use shifts from 2006 to 2019 to construct the instrument. Second, the instrument remains strong and the IV coefficient

[11]Figure IA12 in the Internet Appendix plots the distributions of majors and degrees among inventors and non-inventors, while Table IA12 reports the top 30 university programs with the highest fraction of graduates who become inventors. Together, these patterns indicate that university programs are strong predictors of future inventors.

remains significantly positive when standard errors are clustered at the firm level (Table IA14) or at the city-field-year by firm level (Table IA15). Third, Table IA16 shows an alternative regression where we instead instrument the number of inventors in the same field at the firm. The instrument remains strong and the IV estimate remains significantly positive (0.068, $p < 0.05$), confirming the spillover effects among inventors working on related technologies.

4 Within-City Innovation Spillovers

A large literature has documented that innovation benefits from geographic agglomeration, particularly the concentration of inventors in the same city and field (e.g., Jaffe, Trajtenberg, and Henderson, 1993; Moretti, 2021). In this section, we examine whether city-level spillovers remain important once within-firm spillovers are taken into account.

4.1 City-Level Innovation Spillovers

We first estimate the effect of city-level inventor concentration on individual inventor productivity, without controlling for firm-level inventor counts. We augment equation (1) with city-level measures:

$$y_{i,f,c,k,t} = \gamma_1 \ln(\textit{City_FieldInv}_{c,k,t}) + \gamma_2 \ln(\textit{City_OtherInv}_{c,k,t}) + \mathbf{\Gamma'} \boldsymbol{X}_{i,t} + \alpha_{c,k} + \alpha_{k,t} + \phi_i + \varepsilon_{i,f,c,k,t}, \tag{11}$$

where $\textit{City_FieldInv}_{c,k,t}$ is the number of same-field inventors in city c in year t and $\textit{City_OtherInv}_{c,k,t}$ is the number of other-field inventors. Because the key regressors vary at the city-field-year level, we cluster standard errors at the city-field level in this specification.

Table 7 reports the results. In the baseline specification with city-field and field-year fixed effects, the coefficient on $\ln(\textit{City_FieldInv})$ is 0.086 ($p < 0.05$). Adding inventor fixed effects increases the estimate to 0.113 ($p < 0.01$), suggesting that the cross-sectional correlation is not inflated by the sorting of more productive inventors into larger cities. Including firm fixed effects in Column (3) yields a coefficient of 0.084 ($p < 0.01$), indicating that city-level same-field inventor density has a robust positive association with inventor productivity, even after controlling for individual and firm heterogeneity. These estimates are consistent with prior findings of geographic agglomeration effects on innovation (e.g., Moretti, 2021). In the next subsection, we

examine whether these city-level effects survive once within-firm inventor composition is directly controlled for. Table IA17 in the Internet Appendix replicates the specification in Moretti (2021) on our data and confirms his finding of a positive city-level effect among star inventors: the coefficient on log city-field inventor density is 0.096 ($p < 0.01$) in the most saturated specification. Table IA18 further examines city-level spillovers for star inventors at or above the ninetieth percentile in total patent count.

4.2 Within-Firm versus Within-City Spillovers

We next estimate a horse-race specification that includes both firm-level and city-level inventor counts:

$$\begin{aligned} y_{i,f,c,k,t} = {} & \beta_1 \ln(\mathit{Firm_FieldInv}_{f,k,t}) + \beta_2 \ln(\mathit{Firm_OtherInv}_{f,t}) + \beta_3 \ln(\mathit{Firm_NonInv}_{f,t}) \\ & + \gamma_1 \ln(\mathit{City_FieldInv}_{c,k,t}) + \gamma_2 \ln(\mathit{City_OtherInv}_{c,t}) + \alpha_{c,k} + \alpha_{k,t} + \phi_i + \varepsilon_{i,f,c,k,t}, \end{aligned} \tag{12}$$

Note that, relative to equation (1), we now decompose the firm-level inventor count into same-field inventors (*Firm_FieldInv*) and other-field inventors (*Firm_OtherInv*), so the coefficients β_1, β_2, and β_3 in equation (12) are not directly comparable to those in equation (1). Importantly, the city-level inventor counts *exclude* inventors employed at the same firm as inventor i, so that the firm and city effects are not mechanically confounded.[12]

Table 8 presents the results. The within-firm spillover—the coefficient on $\ln(\mathit{Firm_FieldInv})$—is 0.062 ($p < 0.01$) in the most saturated specification, similar in magnitude to the field-inventor effect in Table 3. In contrast, the city-level same-field inventor coefficient falls from a positive and significant value without inventor fixed effects to 0.002 (statistically insignificant) once inventor and firm fixed effects are included. This pattern is consistent with the interpretation that city-level spillovers largely reflect the sorting of more productive inventors and firms into technology clusters, rather than independent city-level knowledge flows. Tables IA19 and IA20 in the Internet Appendix confirm that this pattern holds on the intensive margin (with a coefficient of 0.067 on $\ln(\mathit{Firm_FieldInv})$) and among star inventors (coefficient of 0.032).

We acknowledge an alternative interpretation of these results. If city-level agglomeration operates *through* firms—for example, by attracting higher-quality firms to technology clusters or by

[12] That is, $\mathit{City_FieldInv}_{c,k,t}$ counts all same-field inventors in city c in year t except those at firm f, and similarly for $\mathit{City_OtherInv}_{c,t}$.

enabling firms to assemble better inventor teams—then controlling for firm-level characteristics would absorb the very channel through which city spillovers operate. Under this interpretation, the attenuation of city-level coefficients does not mean that geographic agglomeration is unimportant, but rather that its benefits are mediated by the firm. Distinguishing between "no city spillovers" and "city spillovers that operate through firms" is inherently difficult, and our results are consistent with both views. What the results do establish is that, conditional on the firm environment, residual city-level effects on inventor productivity are small.

4.3 Citation Source Decomposition

To further understand the channels through which within-firm spillovers operate, we decompose forward citations by their source: citations from within the same firm, citations from other firms in the same city, and citations from firms outside the city. Figure 6 shows that within-firm citations account for a disproportionate share of total forward citations: on average, a patent receives 1.02 citations from within the same firm, compared with 0.74 from other firms in the same city and 5.49 from firms outside the city. While the absolute number of within-firm citations is smaller than out-of-city citations (reflecting the far larger pool of potential citers outside the city), the within-firm citation rate per potential citer is substantially higher.

Table 9 formalizes this by regressing per-patent citations from each source on the number of firm-level inventors and non-inventors. The positive effect of the number of inventors on within-firm citations (coefficient of 0.274, $p < 0.01$) is large relative to the mean of 1.02, whereas the effect on within-city citations from other firms is smaller (0.140). This is consistent with the interpretation that within-firm knowledge spillovers are the primary mechanism.

4.4 Inventor Spin-offs

Finally, we examine whether within-firm inventor interactions extend to entrepreneurship. Prior work suggests that exposure to innovative colleagues can generate entrepreneurial ideas and skills (e.g., Azoulay, Graff Zivin, and Wang, 2010). If within-firm knowledge spillovers expose inventors to a richer set of commercializable ideas, inventors at firms with more fellow inventors may be more likely to found their own companies.

Table 10 tests this prediction. The dependent variable is an indicator for whether an inventor becomes a founder within one, three, or five years. For the five-year horizon, the coefficient on

$\ln(Firm_Inv)$ is 0.00076 ($p < 0.01$), which is large relative to the base rate of founding (0.52%). The number of non-inventors has the opposite effect: its coefficient is -0.0019 ($p < 0.01$), suggesting that organizational scale discourages entrepreneurial exit while fellow inventors encourage it.

We note that this association does not necessarily reflect spillovers per se. Firms with more inventors may also be more innovative, exposing inventors to a wider range of commercializable technologies; alternatively, large-inventor firms may have weaker retention mechanisms. Table IA21 in the Internet Appendix shows that inventor-led startups substantially outperform other startups: they are 20.5 percentage points more likely to patent within five years (base rate: 6.1%), have 22.6% higher employment at year three, and are 18.0 percentage points more likely to receive venture funding (base rate: 22.7%). These patterns are consistent with the interpretation that within-firm interactions enhance the human capital that founders bring to their ventures.

5 Conclusion

This paper documents a positive relationship between inventor productivity and the number of inventors at the firm, alongside a negative relationship with non-inventor employment. We find that a 10 percent increase in the number of inventors raises individual citation-weighted patent output by 2.2 percent. We bring to bear a mover design, a two-step AKM framework, and a shift-share instrumental variables strategy based on variation in the supply of future inventors across university programs to establish causality. The IV estimates imply that 10 percent more firm inventors increase individual adjusted citations output by 1.8 percent. Furthermore, city-level agglomeration effects largely disappear once firm-level inventor composition is accounted for. Our findings point to the firm as the primary locus of innovation spillovers. The productivity gains from agglomeration appear to operate primarily through interactions within firm boundaries. Estimates of geographic spillovers that do not account for firm structure may overstate the role of cities.

References

Abowd, John M, Francis Kramarz, and David N Margolis (1999). "High wage workers and high wage firms". *Econometrica* 67.2, pp. 251–333.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez (2021). "Automated linking of historical data". *Journal of Economic Literature* 59.3, pp. 865–918.

Abramitzky, Ran, Roy Mill, and Santiago Pérez (2020). "Linking individuals across historical sources: A fully automated approach". *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53.2, pp. 94–111.

Acs, Zoltan J and David B Audretsch (1988). "Innovation in large and small firms: an empirical analysis". *The American Economic Review*, pp. 678–690.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt (2005). "Competition and innovation: An inverted-U relationship". *The Quarterly Journal of Economics* 120.2, pp. 701–728.

Akcigit, Ufuk, Murat Alp Celik, and Jeremy Greenwood (2016). "Buy, Keep, or Sell: Economic Growth and the Market for Ideas". en. *Econometrica* 84.3, pp. 943–984.

Akcigit, Ufuk and Nathan Goldschlag (2025). "Measuring the characteristics and employment dynamics of US inventors". *Journal of Economic Growth*, pp. 1–33.

Akcigit, Ufuk and William R Kerr (2018). "Growth through heterogeneous innovations". *Journal of Political Economy* 126.4, pp. 1374–1443.

Almeida, Paul and Bruce Kogut (1999). "Localization of knowledge and the mobility of engineers in regional networks". *Management science* 45.7, pp. 905–917.

Amanzadeh, Naser, Amir Kermani, and Timothy McQuade (2024). *Return Migration and Human Capital Flows*. Working Paper. UC Berkeley.

Arrow, Kenneth J (1962). "The economic implications of learning by doing". *The review of economic studies* 29.3, pp. 155–173.

Atkin, David, M Keith Chen, and Anton Popov (2022). *The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley*. Tech. rep. National Bureau of Economic Research.

Azoulay, Pierre, Joshua S Graff Zivin, and Jialan Wang (2010). "Superstar extinction". *The Quarterly Journal of Economics* 125.2, pp. 549–589.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel (Jan. 2022). "Quasi-Experimental Shift-Share Research Designs". en. *The Review of Economic Studies* 89.1. Ed. by Dirk Krueger, pp. 181–213. ISSN: 0034-6527, 1467-937X. DOI: 10.1093/restud/rdab030. URL: https://academic.oup.com/restud/article/89/1/181/6294942 (visited on 11/24/2025).

Card, David, Jesse Rothstein, and Moises Yi (Jan. 2025). "Location, Location, Location". en. *American Economic Journal: Applied Economics* 17.1, pp. 297–336.

Chien, Colleen V. and Jillian Grennan (2024). *Unpacking the Innovator-Inventor Gap: Evidence from Engineers*. Tech. rep. 4721522. SSRN. DOI: 10.2139/ssrn.4721522. URL: https://ssrn.com/abstract=4721522.

Darby, Michael R. and Lynne G. Zucker (1999). *California's Science Base: Size, Quality and Productivity*. Sacramento, CA: California Council on Science and Technology.

Edmans, Alex, Vivian W Fang, and Katharina A Lewellen (2017). "Equity vesting and investment". *The Review of Financial Studies* 30.7, pp. 2229–2271.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams (Nov. 2016). "Sources of Geographic Variation in Health Care: Evidence From Patient Migration*". *The Quarterly Journal of Economics* 131.4, pp. 1681–1726.

Giroud, Xavier, Ernest Liu, and Holger Mueller (2024). *Innovation Spillovers across US Tech Clusters*. Tech. rep. National Bureau of Economic Research.

Glaeser, Edward L (1999). "Learning in cities". *Journal of Urban Economics* 46.2, pp. 254–277.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings". *Journal of political economy* 118.3, pp. 536–598.

Griliches, Zvi (1958). "Research costs and social returns: Hybrid corn and related innovations". *Journal of political economy* 66.5, pp. 419–431.

Henderson, Rebecca and Iain Cockburn (1996). "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery". *The RAND Journal of Economics* 27.1, pp. 32–59.

Igami, Mitsuru (2017). "Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998". *Journal of Political Economy* 125.3, pp. 798–847.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations". *The Quarterly journal of Economics* 108.3, pp. 577–598.

Jovanovic, Boyan and Rafael Rob (1989). "The growth and diffusion of knowledge". *The Review of Economic Studies* 56.4, pp. 569–582.

Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Hassan, Josh Lerner, and Ahmed Tahoun (Jan. 2025). "The Diffusion of New Technologies*". *The Quarterly Journal of Economics*, qjaf002. eprint: https://academic.oup.com/qje/advance-article-pdf/doi/10.1093/qje/qjaf002/61489798/qjaf002.pdf.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017). "Technological innovation, resource allocation, and growth". *The Quarterly Journal of Economics* 132.2, pp. 665–712.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury (Apr. 2023). "Do firm effects drift? Evidence from Washington administrative data". *Journal of Econometrics* 233.2, pp. 375–395.

Lerner, Josh and Amit Seru (June 2022). "The Use and Misuse of Patent Data: Issues for Finance and Beyond". *The Review of Financial Studies* 35.6, pp. 2667–2704.

Lucas, Robert E (1988). "On the mechanics of economic development". *Journal of monetary economics* 22.1, pp. 3–42.

Marshall, Alfred (1890). *Principles of economics*. Springer.

Moretti, Enrico (2004a). "Human capital externalities in cities". *Handbook of regional and urban economics*. Vol. 4. Elsevier, pp. 2243–2291.

— (2004b). "Workers' education, spillovers, and productivity: evidence from plant-level production functions". *American Economic Review* 94.3, pp. 656–690.

— (Oct. 2021). "The Effect of High-Tech Clusters on the Productivity of Top Inventors". en. *American Economic Review* 111.10. Publisher: American Economic Association, pp. 3328–3375.

Saxenian, AnnaLee (1996). *Regional advantage: Culture and competition in silicon valley and route 128, with a new preface by the author*. Harvard University Press.

Seru, Amit (2014). "Firm boundaries matter: Evidence from conglomerates and R&D activity". *Journal of financial economics* 111.2, pp. 381–405.

Singh, Jasjit (2005). “Collaborative networks as determinants of knowledge diffusion patterns”. *Management science* 51.5, pp. 756–770.

Zucker, Lynne G. and Michael R. Darby (1999). *California’s Inventive Activity: Patent Indicators of Quantity, Quality, and Organizational Origins*. Sacramento, US: California Council on Science and Technology.

Figure 1: Innovation and Total Firm Employment

This figure plots the relationship between total firm employment and innovation output. The left panel plots patents per employee against total firm employment. The right panel plots patents per inventor against total firm employment. Both panels use binscatters based on firm-year level data.

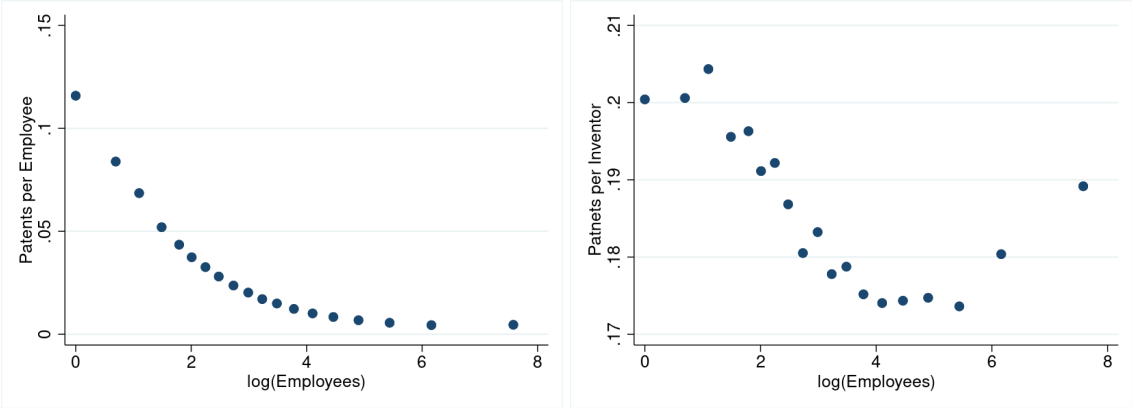


Figure 2: Innovation and Number of Inventors vs. non-inventors

This figure plots the relationship between patents per inventor and the number of non-inventors (left panel) and the number of inventors (right panel) at the firm. The left panel controls for the number of inventors, and the right panel controls for the number of non-inventors. Both panels use binscatters based on firm-year level data.

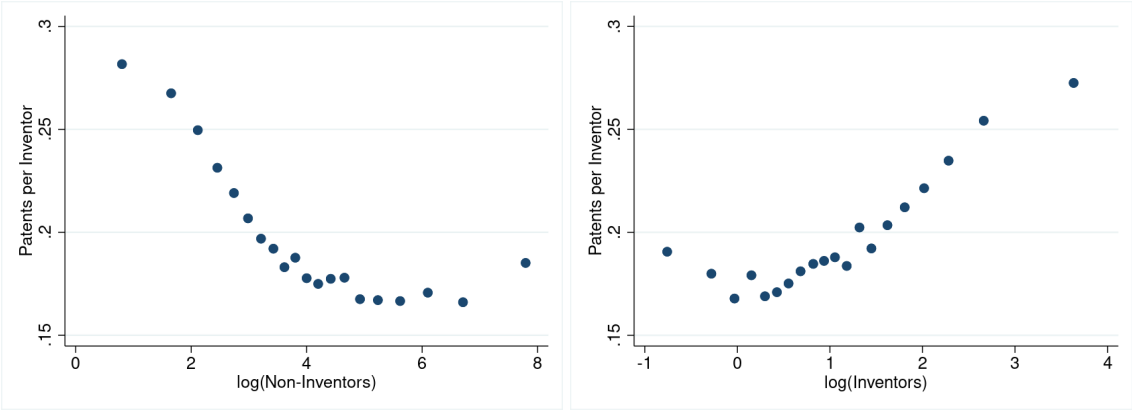


Figure 3: Mover Event Study: Distribution and Event Plot

This figure displays the mover event study following Finkelstein, Gentzkow, and Williams (2016). The left panel shows the distribution of δ_i, defined as the difference in the estimated firm effect between the destination and origin firms. The right panel plots the event study of inventor productivity in the years surrounding the move.

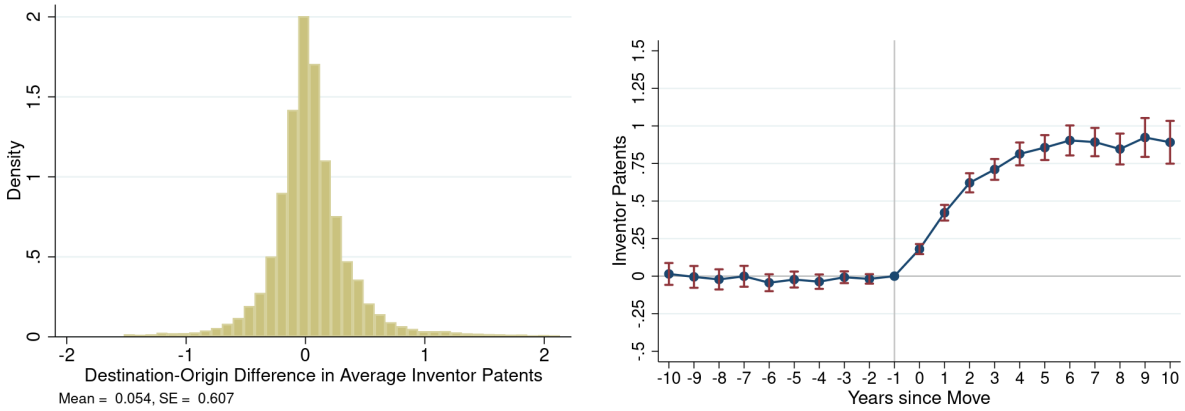


Figure 4: Mover Event Study: Up and Down Moves

This figure plots the event study of inventor productivity separately for inventors moving to a higher-productivity firm (left panel, $\delta_i > 0$) and inventors moving to a lower-productivity firm (right panel, $\delta_i < 0$). δ_i is the difference in the estimated firm effect between the destination and origin firms.

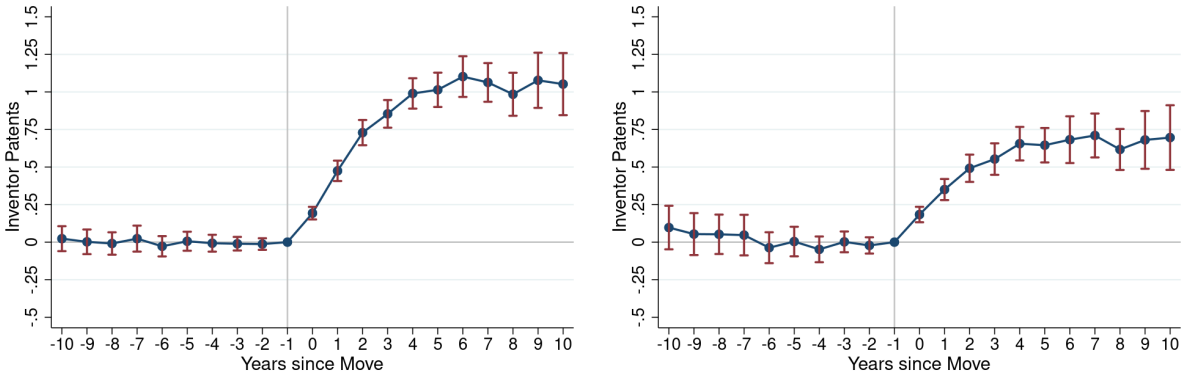


Figure 5: Mover Event Study: Size-based Moves

This figure plots the event study following Equation (6). The left panel shows the distribution of $\delta_{size,i}$, defined as the log difference in the average number of inventors between the destination and origin firms when the mover stayed at the corresponding firms. The right panel plots the event study of inventor productivity in the years surrounding the move.

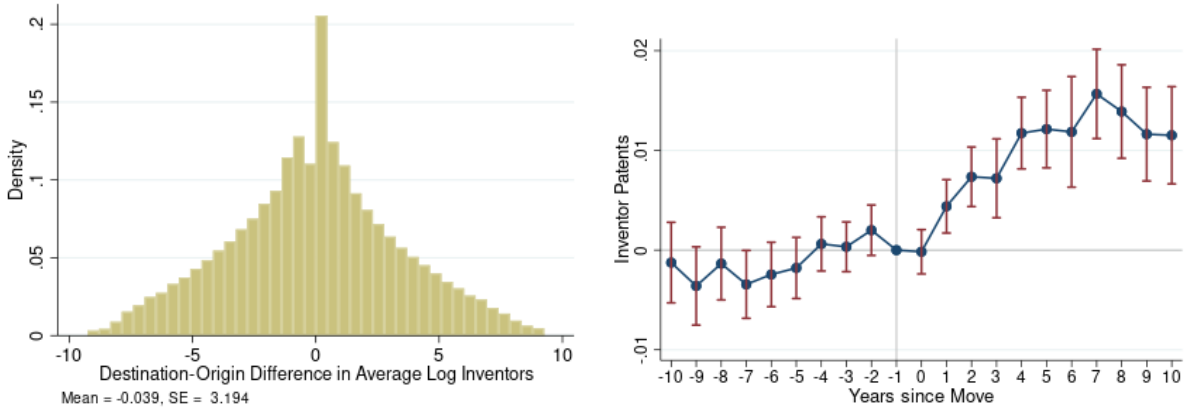


Figure 6: Forward Citation Sources

This figure plots the average number of forward citations received by a patent, decomposed by the source of the citation: from within the same firm, from other firms in the same city, and from firms outside the city.

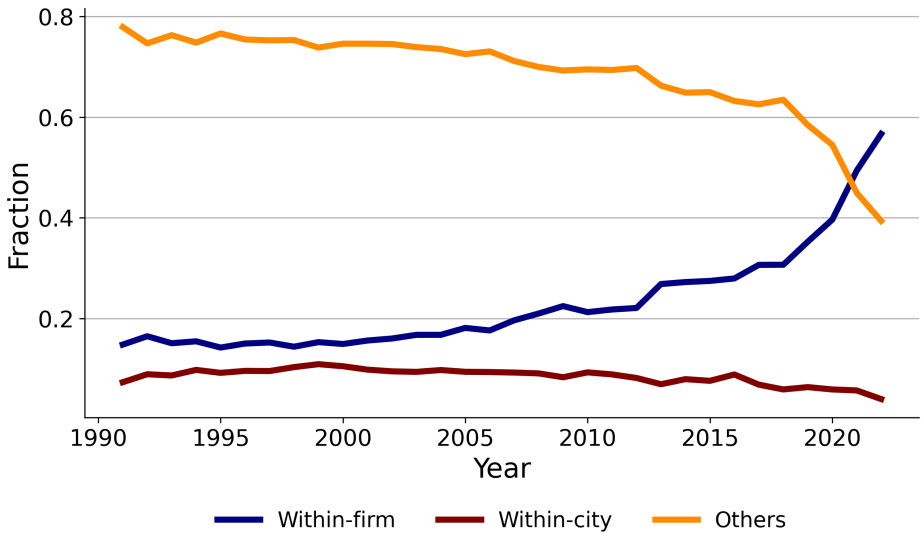


Table 1. Summary Statistics

Panel A reports summary statistics at the inventor-year level for matched inventors from 1979 to 2022. Panel B reports summary statistics at the firm-year level. Panel C reports summary statistics at the city-field-year level, where a city is defined as a Bureau of Economics Area and a field is one of five research fields, both following Moretti (2021).

	Mean	S.D.	25th Pct.	Median	75th Pct.
Panel A: Inventor-year level					
Patent counts	0.257	0.786	0	0	0
Adjusted citations	0.349	1.889	0	0	0
Job experience (years)	12.802	8.532	6	11	18
N	5,938,189				
Panel B: Firm-year level					
Firm_Inv	35.493	194.306	4	7	16
Firm_NonInv	837.564	3925.959	29	101	427
Firm_Emp	873.056	4023.126	36	111	452
N	184,714				
Panel C: City-field-year level					
City_FieldInv	316.794	1140.728	9	33	153
City_OtherInv	1258.906	3527.488	54	192	855
N	28,195				

Table 2. Innovation Spillovers Within Firms: Inventors vs. Non-inventors

This table presents OLS estimates of equation (1). The dependent variable is adjusted citations, defined as the number of citations a patent receives scaled by the average citations for patents in the same year and CPC class. *Firm_Inv* is the number of inventors at the firm. *Firm_NonInv* is the number of non-inventors. Standard errors are clustered at the firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	0.044***	0.058***	0.056***	0.076***
	(0.005)	(0.004)	(0.004)	(0.010)
Ln(Firm_NonInv)	-0.059***	-0.055***	-0.053***	-0.026***
	(0.006)	(0.004)	(0.004)	(0.010)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	5625610	5609706	5606616	5605582
Adj. R^2	0.055	0.271	0.272	0.278
Mean of *y* Variable	0.349			

Table 3. Innovation Spillovers Within Firms: Field Inventor Decomposition

This table decomposes the inventor count into same-field inventors (*Firm_FieldInv*), other-field inventors (*Firm_OtherInv*), and non-inventors (*Firm_NonInv*). The dependent variable is adjusted citations. Standard errors are clustered at the firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_FieldInv)	0.047***	0.061***	0.060***	0.060***
	(0.006)	(0.004)	(0.004)	(0.004)
Ln(Firm_OtherInv)	-0.005	-0.009**	-0.010***	-0.001
	(0.006)	(0.003)	(0.004)	(0.006)
Ln(Firm_NonInv)	-0.054***	-0.048***	-0.046***	-0.011
	(0.006)	(0.004)	(0.004)	(0.010)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	5625610	5609706	5606616	5605582
Adj. R^2	0.056	0.272	0.272	0.278
Mean of *y* Variable	0.346			

Table 4. Innovation Spillovers Within Firms: Stayers vs. Movers

This table examines within-firm spillovers separately for stayers (inventor-years when the inventor remained at the same firm for at least three years) and movers (inventor-years that are three years before and after the inventor switched firms). The dependent variable is adjusted citations. Standard errors are clustered at the firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	Stayers	Movers
	(1)	(2)
Ln(Firm_FieldInv)	0.056***	0.057***
	(0.005)	(0.005)
Ln(Firm_OtherInv)	0.013	-0.003
	(0.009)	(0.008)
Ln(Firm_NonInv)	-0.038**	-0.031**
	(0.016)	(0.013)
Inventor FE	Y	Y
City-field-year FE	Y	Y
Firm FE	Y	Y
N	3142849	1800378
Adj. R^2	0.329	0.292
Mean of y Variable	0.351	0.359

Table 5. Two-Step AKM: Second-Stage Estimates

This table reports second-stage results of the two-step AKM estimation. The first stage estimates equation (2) with inventor, firm-year, and city-field-year fixed effects. The second stage regresses the estimated firm-year effects on log number of inventors and non-inventors at the firm as in equation (3). Standard errors are reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)
Ln(Firm_Inv)	0.037***	0.037***	0.043***
	(0.003)	(0.003)	(0.005)
Ln(Firm_NonInv)	-0.026***	-0.026***	0.008
	(0.003)	(0.003)	(0.006)
Year FE		Y	Y
Firm FE			Y
N	347529	347529	345059
Adj. R^2	0.001	0.002	0.328

Table 6. Innovation Spillovers Within Firms: Instrument Variable

This table presents OLS and IV estimates of equation (10) The dependent variable is adjusted citations, defined as the number of citations a patent receives scaled by the average citations for patents in the same year and CPC class. *Firm_Inv* is the number of inventors at the firm. $\widehat{Firm_Inv}$ is the predicted number of inventors from equation (9). Column (1) presents the OLS estimates, Column (2) presents the IV estimates, and Column (3) presents the first stage estimates. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	OLS	IV	First Stage
	(1)	(2)	(3)
	Adjusted citations	*Adjusted citations*	*Ln(Firm_Inv)*
Ln(Firm_Inv)	0.109***	0.075***	
	(0.015)	(0.025)	
$Ln(\widehat{Firm_Inv})$			0.792***
			(0.026)
City-field-year FE	Y	Y	Y
Inventor FE	Y	Y	Y
Firm FE	Y	Y	Y
N	2521534	2521534	2521534
Adj. R^2	0.363	0.363	0.995
First-stage F		963.932	
Mean of *y* Variable	0.418		

Table 7. Innovation Spillovers Within Cities

This table presents OLS estimates of city-level innovation spillovers. The dependent variable is adjusted citations. *City_FieldInv* is the number of same-field inventors in the city (including those in the same firm). *City_OtherInv* is the number of other-field inventors in the city. Standard errors are clustered at the city-field level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)
Ln(City_FieldInv)	0.086**	0.113***	0.084***
	(0.036)	(0.025)	(0.026)
Ln(City_OtherInv)	0.017	-0.025	-0.057*
	(0.042)	(0.030)	(0.030)
City-field FE	Y	Y	Y
Field-year FE	Y	Y	Y
Inventor FE		Y	Y
Firm FE			Y
N	5624654	5608747	5607711
Adj. R^2	0.055	0.271	0.277
Mean of y Variable	0.349		

Table 8. Innovation Spillovers Within Firms and Within Cities

This table presents the horse-race specification including both firm-level and city-level inventor counts. The dependent variable is adjusted citations. City-level inventor counts *exclude* inventors in the same firm. Standard errors are clustered at the firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)
Ln(Firm_FieldInv)	0.050***	0.062***	0.062***
	(0.006)	(0.004)	(0.004)
Ln(Firm_OtherInv)	-0.005	-0.009**	0.000
	(0.007)	(0.004)	(0.006)
Ln(Firm_NonInv)	-0.055***	-0.048***	-0.019*
	(0.006)	(0.004)	(0.010)
Ln(City_FieldInv)	0.147**	0.041*	0.002
	(0.066)	(0.023)	(0.023)
Ln(City_OtherInv)	-0.037	0.002	-0.028
	(0.052)	(0.029)	(0.028)
City-field FE	Y	Y	Y
Field-year FE	Y	Y	Y
Inventor FE		Y	Y
Firm FE			Y
N	5624014	5608106	5607070
Adj. R^2	0.056	0.272	0.278
Mean of *y* Variable	0.347		

Table 9. Forward Citation Sources and Firm Size

This table decomposes per-patent forward citations by source: from within the same firm, from other firms in the same city, and from firms outside the city. The dependent variable in each column is the per-patent citations from a given source. Standard errors are clustered at the firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	Total	Within-Firm	Within-City	Others
	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	1.277***	0.274***	0.140***	0.827***
	(0.182)	(0.038)	(0.032)	(0.133)
Ln(Firm_NonInv)	2.261***	0.280***	0.285***	1.651***
	(0.329)	(0.052)	(0.074)	(0.215)
City-field-year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
N	358500	358500	358500	358500
Adj. R^2	0.224	0.275	0.176	0.196
Mean of *y* Variable	7.403	1.024	0.739	5.490

Table 10. Inventor Spin-offs

This table examines the relationship between firm inventor composition and the probability that an inventor founds a startup. The dependent variable is an indicator for whether inventor *i* becomes a founder within one year (column 1), three years (column 2), or five years (column 3). Standard errors are clustered at the firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	$Founder_{t+1}$	$Founder_{[t+1,t+3]}$	$Founder_{[t+1,t+5]}$
	(1)	(2)	(3)
Ln(Firm_Inv)	0.00030***	0.00048***	0.00076***
	(0.00007)	(0.00015)	(0.00023)
Ln(Firm_NonInv)	-0.00078***	-0.00091***	-0.00190***
	(0.00009)	(0.00017)	(0.00023)
Year FE	Y	Y	Y
Firm FE	Y	Y	Y
N	6368828	6368828	6368828
Adj. R^2	0.061	0.127	0.156
Mean of *y* Variable	0.0011	0.0033	0.0052

Internet Appendix

Innovation Spillovers Within Firms and Within Cities

Zhao Jin, Amir Kermani, Timothy McQuade, and Minghao Yang

IA-1 Matching Procedure

The matching between the patent inventor and the worker's employment history is done in two parts. In part one, we match the patent assignees with the firm in the employment histories. In part two, we match inventors and workers conditional on the match between assignees and firms.

IA-1.1 Firm Match

A key challenge in firm matching is the lack of standardization in firm names. To minimize false matches, we rely on existing identifiers or crosswalks whenever possible, and only match firms based on names as a last resort. We implement the match in three phases.

In phase one, we use the crosswalk between patent assignees and PERMNOs from Kogan et al. (2017) for public firms, and map PERMNOs to CUSIPs using Compustat. This allows us to match the CUSIPs in the employment history data. For public firms whose stocks are dual-listed in other countries, Compustat CUSIPs capture the US listing, while employment history CUSIPs sometimes correspond to the non-US listing. Therefore, we download both US and non-US CUSIPs for dual-listed stocks from Bloomberg, and manually check these links to enhance the matching.

In phase two, we use Crunchbase as an intermediary because (i) firm names are more consistently documented, (ii) both Crunchbase and PatentsView document the firm's city[2], and (iii) both Crunchbase and the employment history data report firms' official websites, all of which improves match reliability. We standardize firm names in both the patent data and Crunchbase by extending the procedure in Akcigit, Celik, and Greenwood (2016). Specifically, we construct a standardized name that follows Akcigit, Celik, and Greenwood (2016) exactly and an alternative standardized name that additionally strips common suffixes indicating firm type (e.g., "international," "laboratory," "manufacturing"). We use the first to reduce false positives and the second to reduce false negatives. We then match firms in the following order:

1. The firms' state and city match exactly, and the standardized firm names are similar.
2. The firms' state matches exactly, and the standardized firm names are similar.
3. The standardized firm names are highly similar.

In steps 1 and 2, we require both standardized-name variants to have a Jaro-Winkler distance greater than 0.85; in step 3, we require both variants to have a Jaro-Winkler distance greater than 0.95.

In phase three, we directly match patent assignees to firms. Because we do not observe firms' cities in the employment history data, and because firm names are substantially noisier, we follow steps 2 and 3 from phase two but tighten the similarity thresholds: in step 2, we require both standardized name variants to have a Jaro-Winkler distance greater than 0.95; in step 3, we require both variants to have a Jaro-Winkler distance greater than 0.99.

IA-1.2 Inventor Match

With the mapping between patent assignees and firms in the employment history data, we match inventors conditional on the assignee-firm match. Specifically, we require that an inventor work at the patent assignee, its parent, or one of its subsidiaries within a five-year window around the patent application year. We allow for parent-subsidiary links because inventors may report their employer

[2]Here, the firm's city does not correspond to the BEA's "economic area" as in the main text. Instead, it refers to the firm's municipality (e.g., Chicago, Denver, San Francisco).

as either the subsidiary or the parent firm. Because inventors can move frequently, whereas assignees are relatively stable over the patent development cycle, this five-year window serves as a buffer for movers.

Because inventors may record their names differently across data sources, two issues commonly arise: (i) first and last names may be swapped, and (ii) first names may appear in different forms (e.g., Bill vs. William, Tim vs. Timothy, Liz vs. Elizabeth). To address these issues, we implement the inventor match in four steps.

In step one, we require the inventor's first name in the patent record to be similar to the first name in the employment history data, and the inventor's last name in the patent record to be similar to the last name in the employment history data.

In step two, we swap the inventor's first and last names in the patent record. Specifically, we require the inventor's first name in the patent record to be highly similar to the last name in the employment history data, and the inventor's last name in the patent record to be highly similar to the first name in the employment history data.

In step three, we map abbreviated first names to full first names in the employment histories using the crosswalk from Abramitzky, Mill, and Pérez (2020) and Abramitzky et al. (2021). We do not apply the crosswalk to patent names because it relies on the gender of the individual, while patent records do not report gender. We then require the inventor's first name in the patent record to be similar to the mapped first name in the employment history data, and the inventor's last name in the patent record to be similar to the last name in the employment history data.

In step four, we require the first four letters of the first names in the patent record to be highly similar to the first four letters of the first names in the employment history data, and the inventor's last name in the patent record to be highly similar to the last name in the employment history data. We drop matches if an inventor's employment history indicates a Chinese, Japanese, Korean, or Indian origin[3], because the first four letters can be identical across many distinct names in these countries.

Because steps two and four tend to produce more false positives, we use a stricter Jaro-Winkler similarity threshold in these steps. Specifically, we treat two names as a match in steps one and three if the Jaro-Winkler similarity is above 0.90, and in steps two and four if the Jaro-Winkler similarity is above 0.95.

[3]We define the origin country of a worker following Amanzadeh, Kermani, and McQuade (2024).

IA-2 Additional Figures

This section presents additional figures referenced in the main text. Figures IA1–IA3 document the quality of the match between patent data and employment histories across different subsamples. Figure IA4 documents the concentration of inventors across firms. Figures IA5 and IA7 provide supplementary binscatter evidence on the relationship between inventor productivity and firm composition, and Figure IA6 documents the correlation between the number of inventors and non-inventors within firms. Figures IA8–IA11 present additional mover event study results, including raw means and alternative patent measures. Figure IA12 shows the distribution of majors and degrees among inventors and non-inventors.

Figure IA1: Matching Quality: By Application Year and Firm Size

This figure plots the matching rate between the patent data and the employment history data. The left panel shows the matching rate by patent application year. The right panel shows the matching rate by firm size.

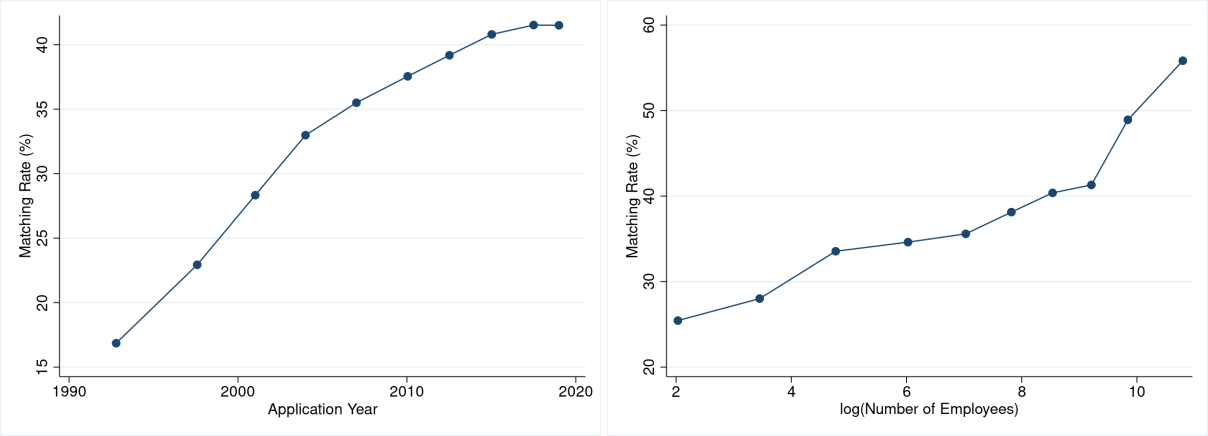


Figure IA2: Matching Quality: By Inventor's Patent Count

This figure plots the matching rate between the patent data and the employment history data by the lifetime patent count of the inventor.

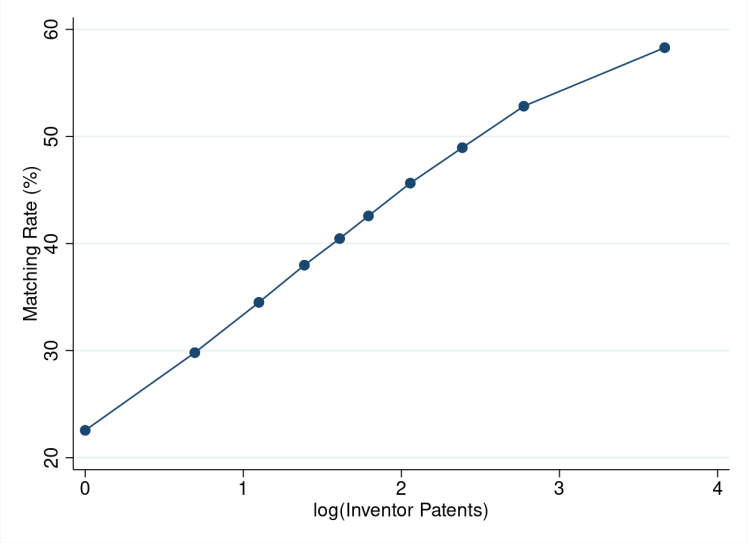


Figure IA3: Conditional Matching Quality

This figure plots the conditional matching rate as a function of an inventor's total patent count. The left panel conditions on inventors with at least two patents. The right panel further conditions on inventors with at least two distinct assignees. The x-axis measures the number of patents of a disambiguated inventor and the y-axis measures the fraction of patents matched to employment histories.

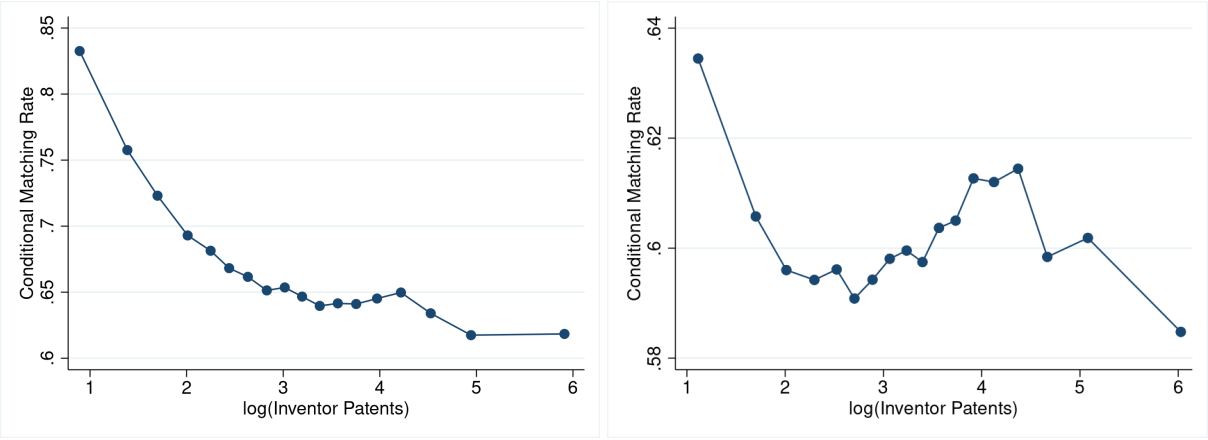


Figure IA4: Inventor Concentration

This figure plots the concentration of inventors in large firms.

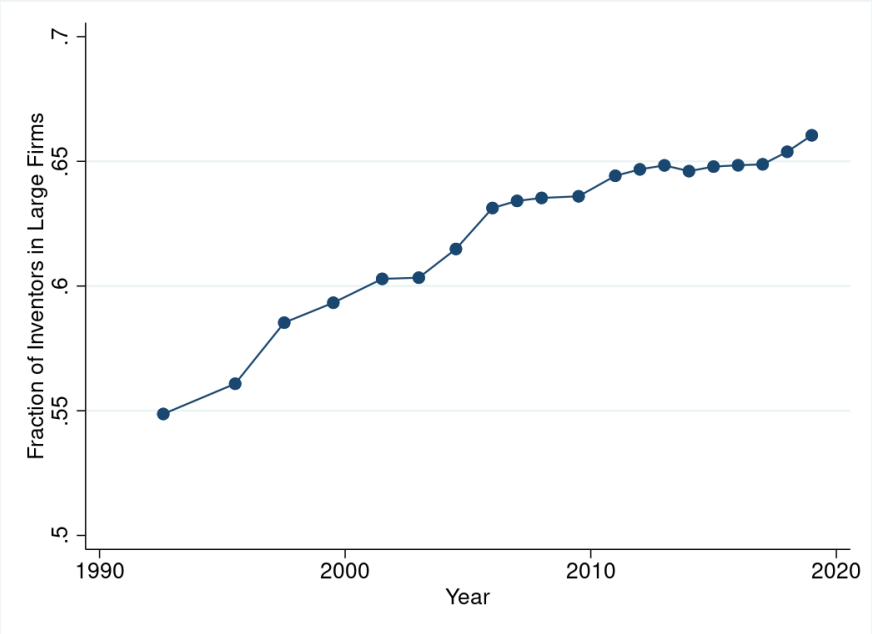


Figure IA5: Innovation and Number of Inventors vs. non-inventors (No Controls)

This figure plots the relationship between patents per inventor and the number of non-inventors (left panel) and the number of inventors (right panel) at the firm, without controlling for the other group.

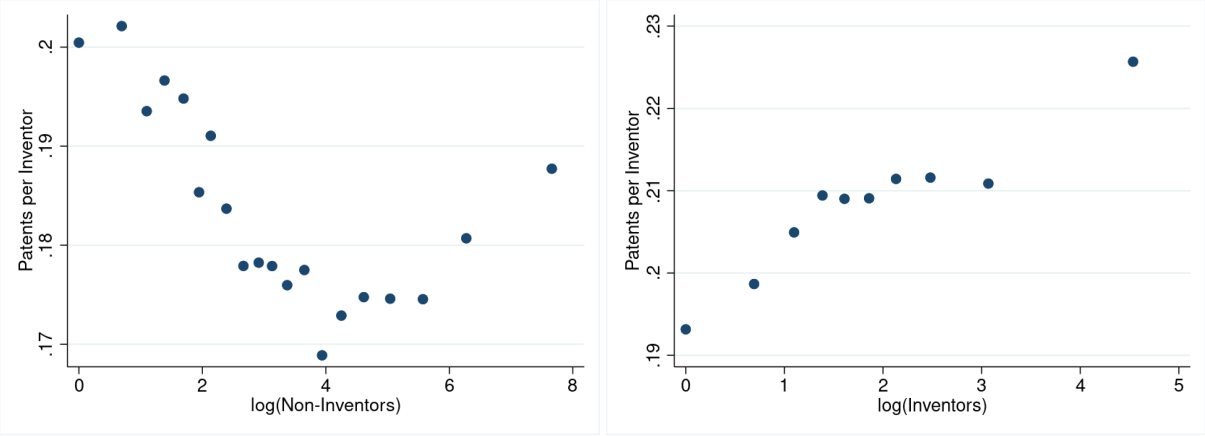


Figure IA6: Correlation Between Number of Inventors and Non-inventors

This figure plots the correlation between the number of inventors and the number of non-inventors at the firm level.

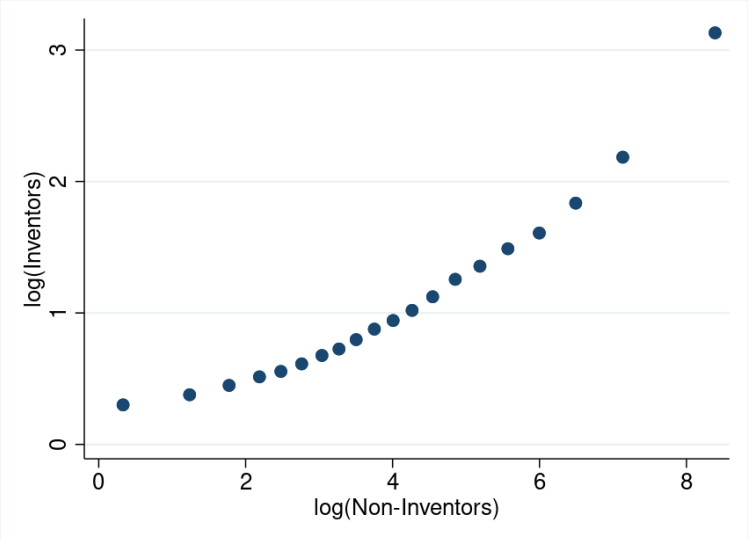
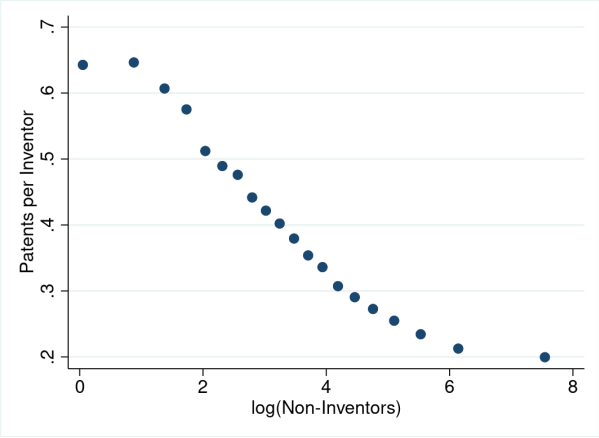


Figure IA7: Innovation and Number of Inventors (PatentsView Inventor Counts)

This figure replicates Figure 2 using inventor counts from PatentsView data instead of matched employment histories.



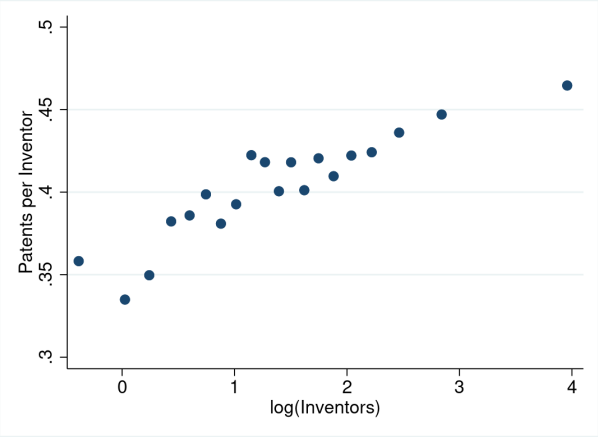


Figure IA8: Mover Event Study: Raw Means

This figure plots the raw mean adjusted citations for movers in the years surrounding the move.

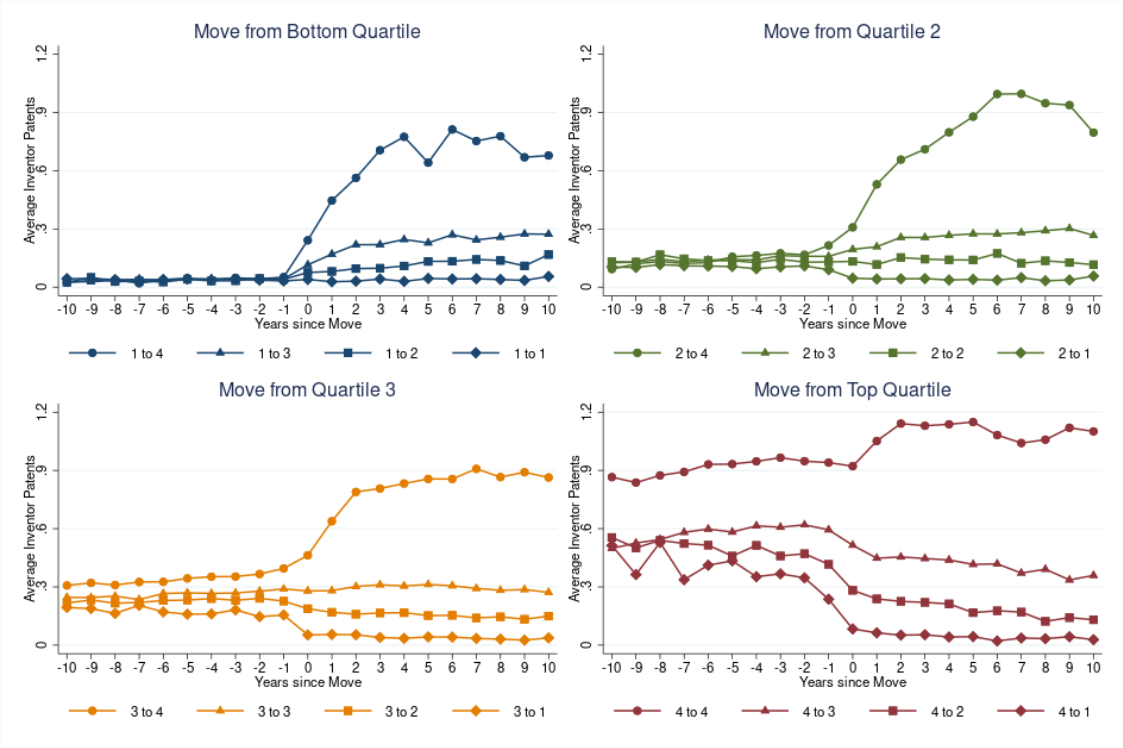


Figure IA9: Event Study: Technology Patents

This figure replicates the event study for technology patents. The left panel shows the distribution of δ_i. The right panel shows the event study plot.

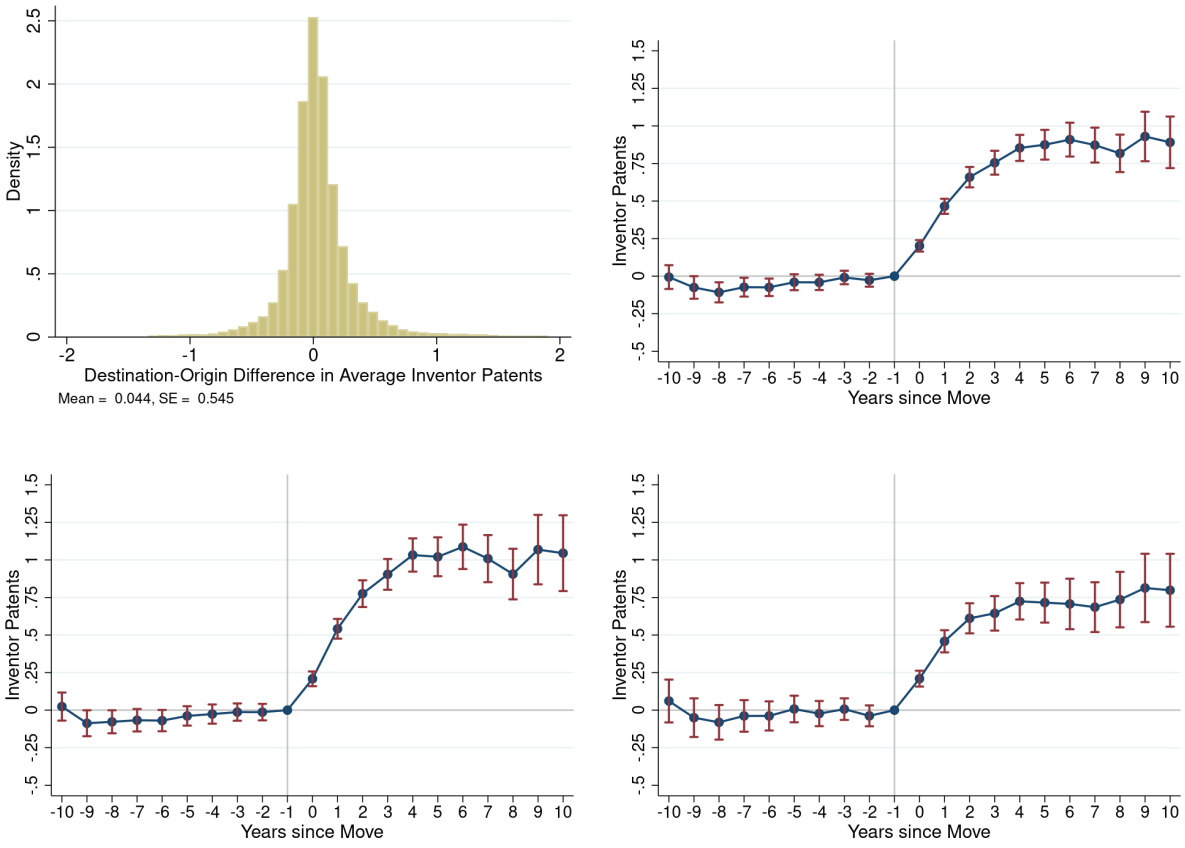


Figure IA10: Event Study: Disruptive Technology Patents

This figure replicates the event study for disruptive technology patents. The left panel shows the distribution of δ_i. The right panel shows the event study plot.

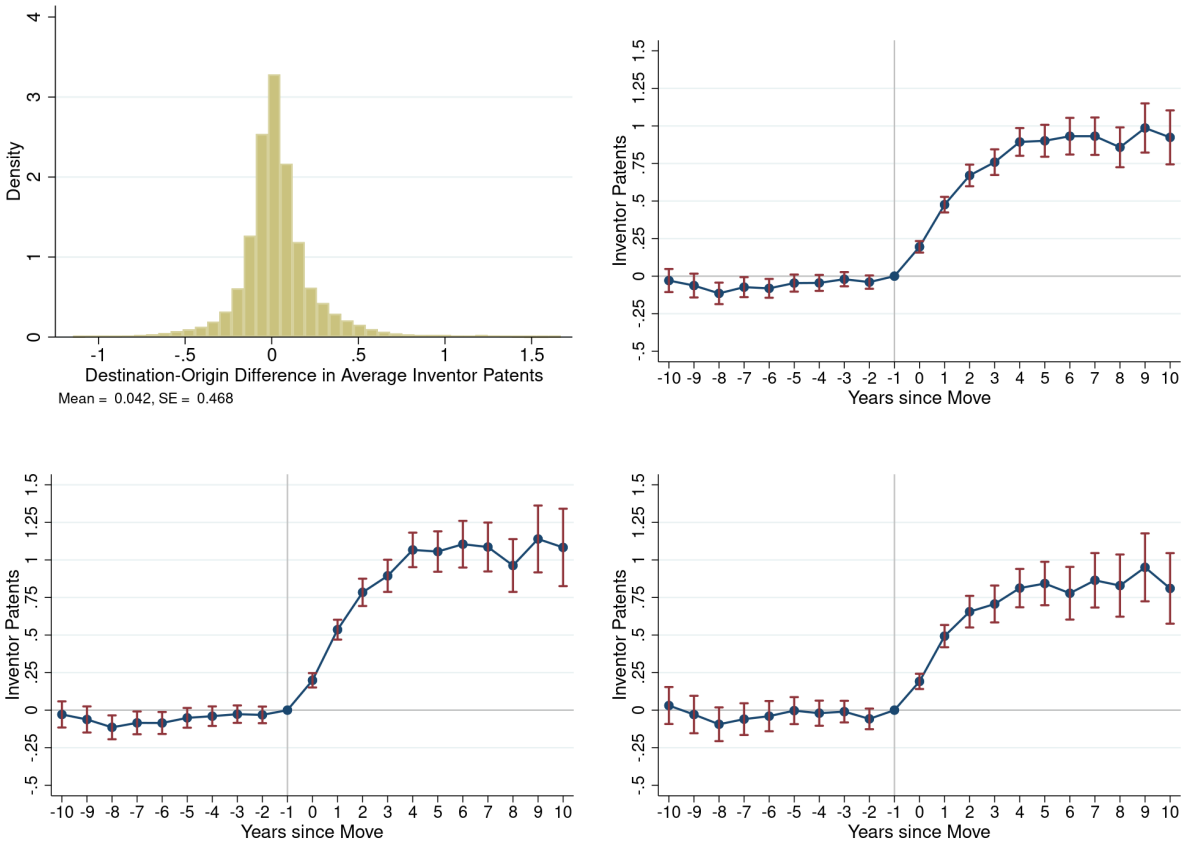
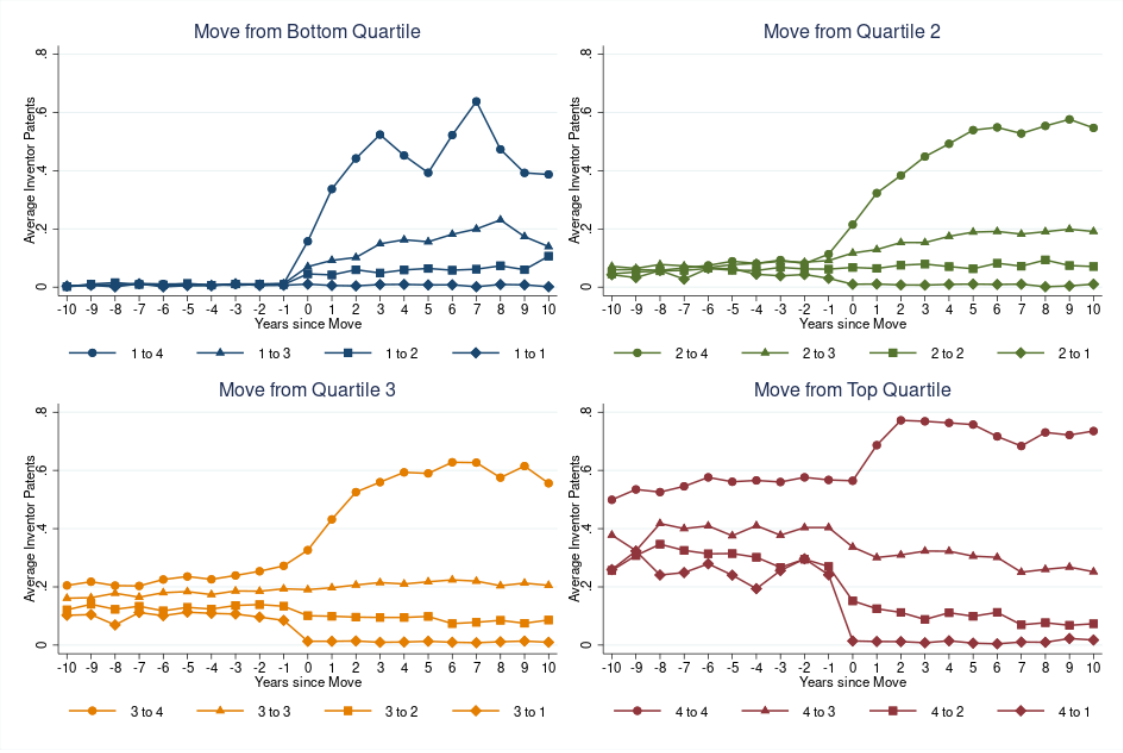


Figure IA11: Event Study Raw Means: Technology and Disruptive Technology Patents

This figure plots the raw mean adjusted citations for movers around the move year, separately for technology patents (top) and disruptive technology patents (bottom).



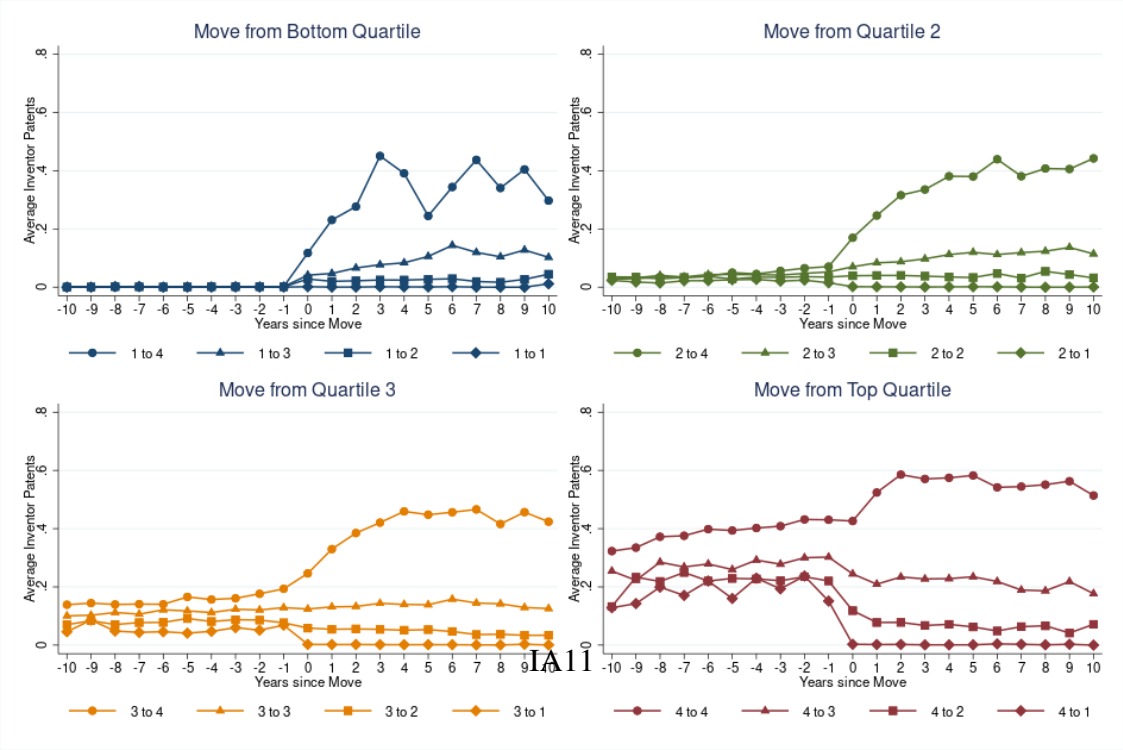
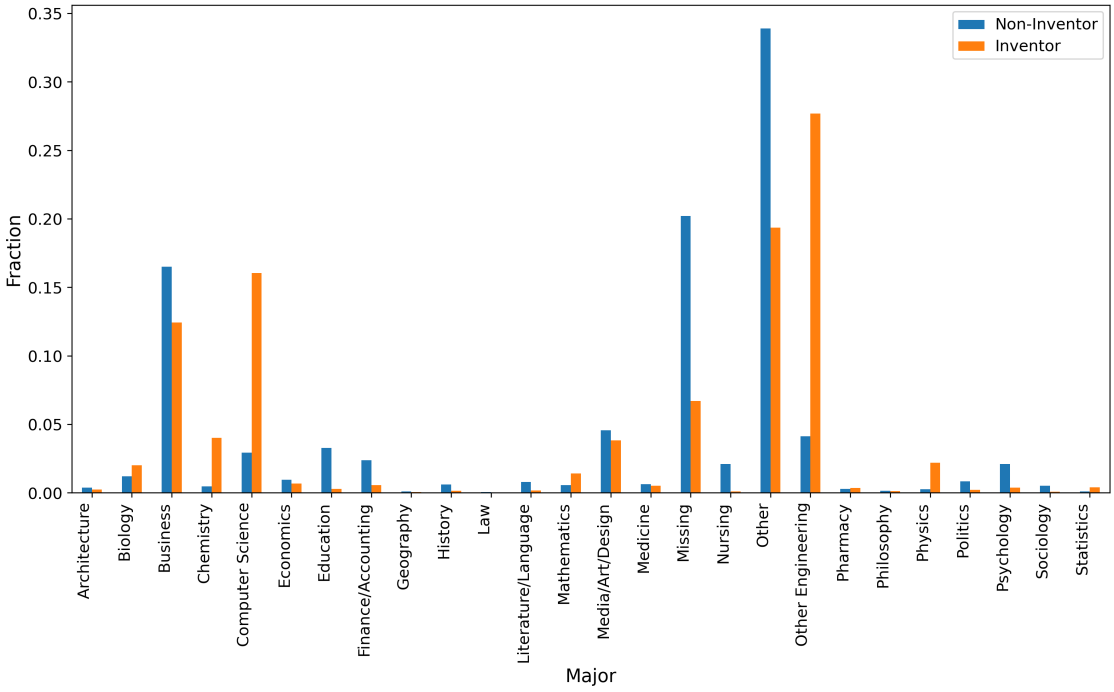
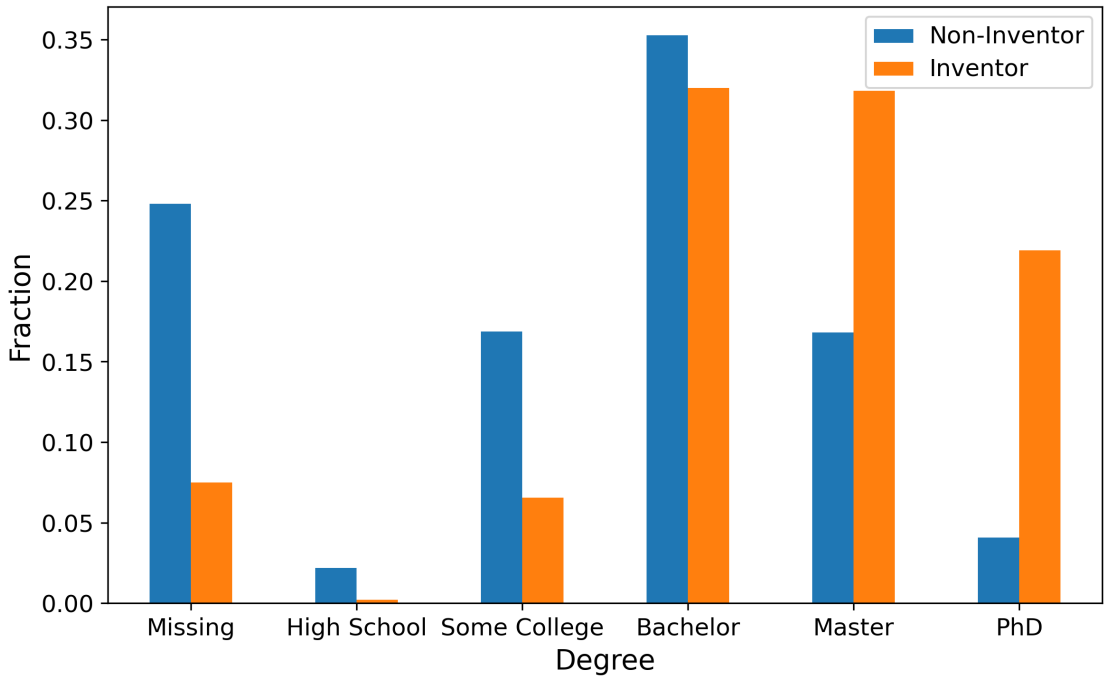


Figure IA12: Major and Degree Distributions of Inventors and Non-Inventors

This figure plots the distribution of inventors and non-inventors across different majors (top) and degrees (bottom).





IA-3 Additional Tables

This section presents additional tables referenced in the main text. Table IA1 reports conditional matching rates between patent data and employment histories. Tables IA2–IA9 report robustness checks for the baseline within-firm spillover estimates using alternative dependent variables (patent counts, log citations), alternative samples (extensive margin, intensive margin, matched patents only), alternative inventor definitions (active versus idle inventors, predicted versus true inventors), and additional controls for public firms. Table IA10 examines heterogeneity in within-firm spillovers. Table IA11 compares the baseline regression results with and without controlling for number of non-inventors within firm. Table IA12 reports the top 30 university programs with the highest shares of graduates becoming inventors in the future. Tables IA13—IA16 report robustness checks for the IV regressions with different share year choice, different standard-error clustering (firm clustering and city-field-year by firm clustering), and different independent variable (number of inventors in the same field at the firm). Tables IA17–IA18 replicate and extend the city-level analysis, including a replication of Moretti (2021) and results for star inventors. Tables IA19–IA20 report robustness checks for the horse-race specification. Table IA21 examines the performance of inventor-led startups.

To construct Table IA8, we collapse the data to the inventor–firm–job level, where a job is defined as a unique combination of job seniority, title, and description. The outcome variable is an inventor indicator equal to one if the employee applies for at least one patent at the firm. The main predictors include the employee's university of graduation, major, degree, job seniority, job title, and job description. We treat university, major, degree, and seniority as categorical variables, and apply TF-IDF to job titles and job descriptions, extracting 2,000 and 8,000 features, respectively. We estimate a logistic regression to obtain predicted probabilities of being an inventor for each inventor–firm–job combination. We then classify combinations with the highest predicted probabilities as inventors until the total number of predicted inventors reaches 6.67 times the number of observed inventors.[4] This procedure yields firm-level inventor and non-inventor counts over time.

[4] Chien and Grennan (2024) document that only 15% of R&D employees ultimately apply for a patent.

Table IA1. Conditional Matching Rates

This table reports the matching rate between the patent data and the employment history data under different conditions. Panel A conditions on the number of employees in the firm. Panel B conditions on the lifetime patent count of an inventor. Panel C conditions on the application year of the patents.

Panel A: By firm size		
$\geq$ 40 Employees	$\geq$ 240 Employees	$\geq$ 1,000 Employees
42.23%	43.11%	44.40%
Panel B: By innovator's patent count		
$\geq$ 2 Patents	$\geq$ 5 Patents	$\geq$ 10 Patents
39.79%	48.37%	53.74%
Panel C: By year of patent		
Before 2000	2000 - 2010	After 2010
16.72%	34.50%	41.70%

Table IA2. Robustness: Patent Counts

This table replicates Table 2 using raw patent counts as the dependent variable instead of adjusted citations. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	0.024***	0.034***	0.032***	0.052***
	(0.002)	(0.002)	(0.002)	(0.005)
Ln(Firm_NonInv)	-0.021***	-0.025***	-0.024***	-0.016**
	(0.002)	(0.002)	(0.002)	(0.006)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	5625610	5609706	5606616	5605582
Adj. R^2	0.165	0.370	0.372	0.375
Mean of y Variable	0.257			

Table IA3. Robustness: Log Adjusted Citations

This table replicates Table 2 using ln(adjusted citations) as the dependent variable. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	0.002	0.023***	0.022***	0.048***
	(0.009)	(0.007)	(0.007)	(0.017)
Ln(Firm_NonInv)	-0.049***	-0.043***	-0.039***	-0.098***
	(0.008)	(0.006)	(0.006)	(0.014)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	1086526	956967	952357	949957
Adj. R^2	0.098	0.302	0.307	0.313
Mean of y Variable	-0.516			

Table IA4. Robustness: Extensive Margin

This table replicates Table 2 but looking at the extensive margin—the dependent variable is changed to an indicator variable that equals one if adjusted citations are greater than zero and zero otherwise. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	0.041***	0.064***	0.060***	0.095***
	(0.008)	(0.007)	(0.006)	(0.015)
Ln(Firm_NonInv)	-0.070***	-0.060***	-0.057***	-0.015
	(0.008)	(0.007)	(0.007)	(0.015)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	4263453	4236274	4232977	4231443
Adj. R^2	0.055	0.276	0.277	0.284
Mean of *y* Variable	0.349			

Table IA5. Robustness: Intensive Margin

This table replicates Table 2 restricting the sample to the intensive margin—inventor-firms such that the inventor applies for at least one patent at the firm. The dependent variable is adjusted citations. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	0.041***	0.064***	0.060***	0.095***
	(0.008)	(0.007)	(0.006)	(0.015)
Ln(Firm_NonInv)	-0.070***	-0.060***	-0.057***	-0.015
	(0.008)	(0.007)	(0.007)	(0.015)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	4263453	4236274	4232977	4231443
Adj. R^2	0.055	0.276	0.277	0.284
Mean of y Variable	0.349			

Table IA6. Robustness: Matched Patents Only

This table replicates Table 2 counting only patents that are matched to employment histories in the dependent variable. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	0.049***	0.078***	0.075***	0.104***
	(0.004)	(0.004)	(0.004)	(0.010)
Ln(Firm_NonInv)	-0.066***	-0.073***	-0.071***	-0.040***
	(0.005)	(0.004)	(0.004)	(0.010)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	5437035	5423301	5420245	5419275
Adj. R^2	0.046	0.234	0.234	0.243
Mean of y Variable	0.227			

Table IA7. Robustness: Active vs. Idle Inventors

This table decomposes the inventor count into active inventors (those who applied for at least one patent in the past five years) and idle inventors (those who did not). The dependent variable is adjusted citations. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_ActiveInv)	0.026**	0.053***	0.050***	0.055***
	(0.011)	(0.004)	(0.004)	(0.005)
Ln(Firm_IdleInv)	0.016	0.002	0.002	0.026***
	(0.011)	(0.005)	(0.005)	(0.007)
Ln(Firm_NonInv)	-0.058***	-0.054***	-0.052***	-0.035***
	(0.006)	(0.004)	(0.004)	(0.010)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	5521885	5505231	5502061	5500739
Adj. R^2	0.056	0.273	0.273	0.279
Mean of *y* Variable	0.349			

Table IA8. Robustness: Predicted Inventors

This table predicts the inventor and non-inventor counts within firm using education history and job information of the employees. In column (1), the independent variables are true inventor and non-inventor counts; in column (2), the independent variables are predicted inventor and non-inventor counts, where employees' education and job seniority are used as features for prediction; in column (3), the independent variables are predicted inventor and non-inventor counts, where employee's education, job seniority, job title, and job description are used as features for prediction. The dependent variable is adjusted citations. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	Actual Inventors	*Predicted Inventors*	*Predicted Inventors*
	(1)	(2)	(3)
x1	0.077***	0.032**	0.040***
	(0.011)	(0.014)	(0.014)
x2	-0.017	0.017	0.009
	(0.012)	(0.012)	(0.011)
Inventor FE	Y	Y	Y
City-field-year FE	Y	Y	Y
Firm FE	Y	Y	Y
Inputs for Prediction:			
Education		Y	Y
Job Seniority		Y	Y
Job Title and Description			Y
N	4838116	4838116	4838116
Adj. R^2	0.291	0.291	0.291
Mean of y Variable	0.374		

Table IA9. Robustness: Public Firms and Additional Controls

This table restricts the sample to public firms, and add additional controls in the regressions. Column (1) does not add any additional controls; column (2) controls for firm age square; column (3) controls for research and development expenditure; column (4) controls for selling, general, and admin expenditure. The dependent variable is adjusted citations. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_FieldInv)	0.0553***	0.0554***	0.0554***	0.0574***
	(0.0066)	(0.0067)	(0.0074)	(0.0067)
Ln(Firm_OtherInv)	0.0067	0.0066	0.0062	0.0103
	(0.0099)	(0.0101)	(0.0120)	(0.0101)
Ln(Firm_NonInv)	-0.0222	-0.0319*	-0.0258	-0.0266
	(0.0178)	(0.0173)	(0.0188)	(0.0179)
Firm age^2		-0.0001***		
		(0.0000)		
Ln(R&D)			0.0036	
			(0.0059)	
Ln(SGA)				0.0011
				(0.0070)
Inventor FE	Y	Y	Y	Y
City-field-year FE	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
N	3193259	3193259	2913347	2996150
Adj. R^2	0.291	0.291	0.297	0.294
First-stage F				
Mean of *y* Variable	0.353	0.353	0.363	0.358

Table IA10. Heterogeneity Analysis

This table examines heterogeneity in within-firm spillovers along various dimensions. *All* refers to the full sample. *Large firms* restricts the sample to large firms (firms with more than 30 inventors). *Star inventors* restricts the sample to star inventors (inventors above the ninetieth percentile in total number of patents). The dependent variable is adjusted citations. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	All	Large firms	Star inventors
	(1)	(2)	(3)
Ln(Firm_Inv)	0.076***	0.107***	0.157***
	(0.010)	(0.017)	(0.029)
Ln(Firm_NonInv)	-0.026***	-0.059***	-0.060**
	(0.010)	(0.019)	(0.029)
Inventor FE	Y	Y	Y
City-field-year FE	Y	Y	Y
Firm FE	Y	Y	Y
N	5605582	4589274	1594656
Adj. R^2	0.278	0.280	0.285
Mean of *y* Variable	0.349	0.349	0.952

Table IA11. Innovation Spillovers Within Firms: Employment Composition

This table presents OLS estimates of equations (1) and (10). The dependent variable is adjusted citations, defined as the number of citations a patent receives scaled by the average citations for patents in the same year and CPC class. *Firm_Inv* is the number of inventors at the firm. *Firm_NonInv* is the number of non-inventors. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)
Ln(Firm_Inv)	0.007***	0.056***	0.056***	0.076***
	(0.002)	(0.004)	(0.006)	(0.010)
Ln(Firm_NonInv)		-0.053***		-0.026***
		(0.004)		(0.010)
City-field FE	Y	Y		
Field-year FE	Y	Y		
Inventor FE		Y	Y	Y
City-field-year FE			Y	Y
Firm FE				Y
N	5606616	5606616	5605582	5605582
Adj. R^2	0.272	0.272	0.278	0.278
Mean of *y* Variable	0.349			

Table IA12. Top 30 University Programs in Inventor Shares

This table shows the top 30 university programs with the highest inventor shares from 2005 to 2019, where inventor share is defined as the fraction of graduates becoming an inventor.

University	Major	Degree	N. Graduates	Inv. Share (%)
Scripps College	Chemistry	PhD	197	38.6
Gies College of Business University of Illinois Urbana Champaign	Computer Science	PhD	182	37.9
University of Southern Mississippi	Other Engineering	PhD	88	37.5
Boston College	Chemistry	PhD	148	33.1
University of California Berkeley	Computer Science	PhD	616	32.8
Carnegie Mellon University	Computer Science	PhD	654	31.3
Oregon State University	Computer Science	PhD	116	31.0
University of North Carolina at Chapel Hill	Computer Science	PhD	168	31.0
Massachusetts Institute of Technology	Computer Science	PhD	667	30.7
California Institute of Technology	Computer Science	PhD	109	29.4
Purdue University West Lafayette	Computer Science	PhD	728	29.3
University of California Santa Barbara	Other Engineering	PhD	389	29.0
Stanford University	Computer Science	PhD	479	29.0
University of California Los Angeles	Computer Science	PhD	340	28.8
Harvard University	Chemistry	PhD	354	28.8
Brown University	Computer Science	PhD	122	28.7
Stanford University	Chemistry	PhD	438	28.3
Johns Hopkins University	Computer Science	PhD	225	28.0
University of Forestry	Computer Science	PhD	339	27.7
Georgia Institute of Technology	Computer Science	PhD	971	27.4
Penn State Main Campus	Computer Science	PhD	275	27.3
Stanford University	Other Engineering	PhD	1943	27.2
Gies College of Business University of Illinois Urbana Champaign	Chemistry	PhD	166	27.1
Binghamton University State University of New York	Computer Science	PhD	85	27.1
University of Rochester	Computer Science	PhD	105	26.7
University of Texas at Dallas	Other Engineering	PhD	304	26.6
University of Washington Seattle	Computer Science	PhD	259	26.6
Northwestern University	Computer Science	PhD	207	26.6
California Institute of Technology	Chemistry	PhD	414	26.3
University of Maryland College Park	Computer Science	PhD	560	25.9

Table IA13. Innovation Spillovers Within Firms: Instrument Variable With 2005 Hiring Shares

This table presents OLS and IV estimates of equation 10. The dependent variable is adjusted citations, defined as the number of citations a patent receives scaled by the average citations for patents in the same year and CPC class. *Firm_Inv* is the number of inventors at the firm. $\widehat{Firm_Inv}$ is the predicted number of inventors from equation (9), where we use the firm's hiring share from 2005 instead of 2009. Column (1) presents the OLS estimates, Column (2) presents the IV estimates, and Column (3) presents the first stage estimates. Standard errors are clustered at the city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	OLS	IV	First Stage
	(1)	(2)	(3)
	Adjusted citations	*Adjusted citations*	*Ln(Firm_Inv)*
Ln(Firm_Inv)	0.109***	0.075***	
	(0.015)	(0.025)	
$Ln(\widehat{Firm_Inv})$			0.792***
			(0.026)
City-field-year FE	Y	Y	Y
Inventor FE	Y	Y	Y
Firm FE	Y	Y	Y
N	2521534	2521534	2521534
Adj. R^2	0.363	0.363	0.995
First-stage F		963.932	
Mean of *y* Variable	0.418		

Table IA14. Innovation Spillovers Within Firms: Instrument Variable With Firm Clustering

This table presents OLS and IV estimates of equation (10). The dependent variable is adjusted citations, defined as the number of citations a patent receives scaled by the average citations for patents in the same year and CPC class. *Firm_Inv* is the number of inventors at the firm. $\widehat{Firm_Inv}$ is the predicted number of inventors from equation (9). Column (1) presents the OLS estimates, Column (2) presents the IV estimates, and Column (3) presents the first stage estimates. Standard errors are clustered at the firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	OLS	IV	First Stage
	(1)	(2)	(3)
	Adjusted citations	*Adjusted citations*	*Ln(Firm_Inv)*
Ln(Firm_Inv)	0.109***	0.075**	
	(0.017)	(0.034)	
$Ln(\widehat{Firm_Inv})$			0.792***
			(0.016)
City-field-year FE	Y	Y	Y
Inventor FE	Y	Y	Y
Firm FE	Y	Y	Y
N	2521534	2521534	2521534
Adj. R^2	0.363	0.363	0.995
First-stage F		2539.120	
Mean of *y* Variable	0.418		

Table IA15. Innovation Spillovers Within Firms: Instrument Variable With City-Field-Year By Firm Clustering

This table presents OLS and IV estimates of equation (10). The dependent variable is adjusted citations, defined as the number of citations a patent receives scaled by the average citations for patents in the same year and CPC class. *Firm_Inv* is the number of inventors at the firm. $\widehat{Firm_Inv}$ is the predicted number of inventors from equation (9). Column (1) presents the OLS estimates, Column (2) presents the IV estimates, and Column (3) presents the first stage estimates. Standard errors are clustered at the city-field-year by firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	OLS	IV	First Stage
	(1)	(2)	(3)
	Adjusted citations	*Adjusted citations*	*Ln(Firm_Inv)*
Ln(Firm_Inv)	0.109***	0.075**	
	(0.017)	(0.034)	
$Ln(\widehat{Firm_Inv})$			0.792***
			(0.017)
City-field-year FE	Y	Y	Y
Inventor FE	Y	Y	Y
Firm FE	Y	Y	Y
N	2521534	2521534	2521534
Adj. R^2	0.363	0.363	0.995
First-stage F		2100.526	
Mean of *y* Variable	0.418		

Table IA16. Innovation Spillovers Within Firms: Instrument Variable For Field Inventors

This table presents OLS and IV estimates of equation (10), changing the key independent variable as *Ln(Firm_FieldInv)*. The dependent variable is adjusted citations, defined as the number of citations a patent receives scaled by the average citations for patents in the same year and CPC class. *Firm_FieldInv* is the number of inventors in the same field at the firm. $\widehat{Firm_Inv}$ is the predicted number of inventors from equation (9). Column (1) presents the OLS estimates, Column (2) presents the IV estimates, and Column (3) presents the first stage estimates. Standard errors are clustered at the city-field-year by firm level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	OLS	IV	First Stage
	(1)	(2)	(3)
	Adjusted citations	*Adjusted citations*	*Ln(Firm_FieldInv)*
Ln(Firm_FieldInv)	0.068***	0.068**	
	(0.008)	(0.026)	
$Ln(\widehat{Firm_Inv})$			0.773***
			(0.034)
City-field-year FE	Y	Y	Y
Inventor FE	Y	Y	Y
Firm FE	Y	Y	Y
N	2447146	2447146	2447146
Adj. R^2	0.364	0.364	0.972
First-stage F		527.195	
Mean of *y* Variable	0.418		

Table IA17. Replication of Moretti (2021)

This table replicates the specification in Moretti (2021). The dependent variable is ln(patent counts). *City_FieldInv_Density* is the ratio of the number of inventors in the same city and field to the total number of inventors in the same field. The sample is restricted to star inventors at or above the ninetieth percentile in total patent count. Standard errors are clustered at city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(City_FieldInv_Density)	-0.008	0.063***	0.065***	0.061***	0.041**	0.032*	0.095***	0.096***
	(0.010)	(0.024)	(0.022)	(0.019)	(0.017)	(0.019)	(0.027)	(0.028)
Year	Y	Y	Y	Y	Y	Y	Y	Y
City	Y	Y	Y	Y	Y	Y	Y	Y
Field	Y	Y	Y	Y	Y	Y	Y	Y
Class	Y	Y	Y	Y	Y	Y	Y	Y
City × field		Y	Y	Y	Y	Y	Y	Y
City × class			Y	Y	Y	Y	Y	Y
Field × year				Y	Y	Y	Y	Y
Class × year					Y	Y	Y	Y
Inventor						Y	Y	Y
City × year							Y	Y
Firm								Y
N	834154	834128	831667	831667	831364	826535	825808	822412
Adj. R^2	0.022	0.025	0.032	0.034	0.036	0.254	0.256	0.259
Mean of Dep. Variable	0.025							

Table IA18. City-Level Spillovers: Star Inventors

This table examines city-level spillovers for star inventors (those at or above the ninetieth percentile in total patent count). The dependent variable is ln(patent counts). Standard errors are clustered at city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)
Ln(City_FieldInv)	0.027	0.052**	0.051**
	(0.031)	(0.025)	(0.025)
Ln(City_OutInv)	-0.016	-0.066**	-0.068*
	(0.031)	(0.030)	(0.035)
City-field FE	Y	Y	Y
Field-year FE	Y	Y	Y
Inventor FE		Y	Y
Firm FE			Y
N	834183	829439	826087
Adj. R^2	0.089	0.249	0.252
Mean of y Variable	0.025		

Table IA19. Within-Firm and Within-City Spillovers: Intensive Margin

This table replicates Table 8 restricting the sample to the intensive margin—inventor-firms such that the inventor applies for at least one patent at the firm. The dependent variable is adjusted citations. City-level inventor counts exclude inventors in the same firm. Standard errors are clustered at city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)
Ln(Firm_FieldInv)	0.047***	0.060***	0.067***
	(0.008)	(0.005)	(0.006)
Ln(Firm_OtherInv)	-0.006	-0.010**	0.005
	(0.008)	(0.005)	(0.008)
Ln(Firm_NonInv)	-0.063***	-0.045***	-0.008
	(0.007)	(0.006)	(0.014)
Ln(City_FieldInv)	0.187**	0.040	0.001
	(0.085)	(0.028)	(0.028)
Ln(City_OtherInv)	-0.054	-0.025	-0.048
	(0.063)	(0.035)	(0.034)
City-field FE	Y	Y	Y
Field-year FE	Y	Y	Y
Inventor FE		Y	Y
Firm FE			Y
N	4624143	4609390	4608386
Adj. R^2	0.056	0.282	0.290
Mean of *y* Variable	0.433		

Table IA20. Within-Firm and Within-City Spillovers: Star Inventors

This table replicates Table 8 restricting the sample to star inventors at or above the ninetieth percentile in total patent count. The dependent variable is ln(adjusted citations). City-level inventor counts exclude inventors in the same firm. Standard errors are clustered at city-field by field-year level and reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	(1)	(2)	(3)
Ln(Firm_FieldInv)	0.023***	0.025***	0.032***
	(0.008)	(0.004)	(0.006)
Ln(Firm_OtherInv)	-0.021	-0.007	0.007
	(0.013)	(0.006)	(0.011)
Ln(Firm_NonInv)	-0.044***	-0.037***	-0.098***
	(0.009)	(0.007)	(0.016)
Ln(City_FieldInv)	0.085	-0.094**	-0.126***
	(0.095)	(0.038)	(0.041)
Ln(City_OtherInv)	-0.090	-0.114**	-0.133***
	(0.076)	(0.049)	(0.047)
City-field FE	Y	Y	Y
Field-year FE	Y	Y	Y
Inventor FE		Y	Y
Firm FE			Y
N	632930	624777	622884
Adj. R^2	0.059	0.263	0.269
Mean of y Variable	-0.178		

Table IA21. Inventor-Led Firm Performance

This table examines the performance of startups founded by inventors who spun off from existing firms, relative to other startups. The dependent variables measure innovation output (column 1), employment (column 2), funding (columns 3–4), and exit outcomes (columns 5–6). Standard errors are reported in parentheses. *, **, or *** indicate significance at the 10%, 5%, or 1% level, respectively.

	$Patent_{[t,t+5]}$	$Ln(Emp_{t+3})$	$Funded_{[t,t+3]}$	$Ln(1+Raised_{[t,t+3]})$	$Acq_{[t,t+5]}$	$IPO_{[t,t+10]}$
	(1)	(2)	(3)	(4)	(5)	(6)
Inventor_Spinoff	0.205***	0.226***	0.180***	0.782***	0.035***	0.017***
	(0.009)	(0.023)	(0.009)	(0.034)	(0.005)	(0.003)
Controls	Y	Y	Y	Y	Y	Y
Industry × Est. Year FE	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Team Size FE	Y	Y	Y	Y	Y	Y
N	66253	66253	66253	66253	66253	66253
Adj. R^2	0.139	0.074	0.187	0.201	0.026	0.047
Mean of *y* Variable	0.061	1.547	0.227	0.527	0.036	0.009