Reveal or Conceal?

Employer Learning in the Labor Market for Computer Scientists

Alice H. Wu*

January 26, 2025

Abstract

How does employer learning affect the allocation of talent in the market for research scientists? I study this question using the job histories of 40,000 Ph.D.'s in computer science (CS) matched to their scientific publications and patent applications. Authorship of a CS conference proceeding doubles the probability that a researcher moves to one of the top tech firms in the following year, controlling for her origin firm and experience, implying a strong role for public learning in the matching process between more productive workers and more productive firms. Many higher-quality papers are accompanied by a related patent application, but the application is private information for 18 months. Authors of such papers are somewhat *less* likely to move up the firm ladder in the following year, but are more likely to end up at a top firm within three years, as predicted by a model of employer wage setting with asymmetric information. I estimate a structural version of the model and find that if employers did not learn about workers from post-PhD research, there would be 16% fewer scientific publications by early-career computer scientists. Disclosing patent applications one year faster would increase innovation by 1%, driven by a faster rate of positive assortative matching.

1   Introduction

Identifying talent is critical to the efficient allocation of labor in an economy. A large body of research suggests that workers' abilities are only partially revealed prior to labor market entry, and that substantial learning by employers occurs over the first decade or so of work (e.g., Altonji and Pierret 2001; Farber and Gibbons 1996; Pallais 2014). Existing tests of employer learning, however, rely on only indirect correlates of worker abilities (Kahn 2013; Lange 2007; Schönberg 2007). In most settings researchers cannot see the public signals about worker ability that are assumed to be available to employers in standard learning models, let alone the private signals that only their current employer can see in models of asymmetric learning (Acemoglu and Pischke 1998; Li 2013). The missing data challenge also makes it difficult to quantify the role of employer learning in the reallocation of talent and the efficiency of the process as emphasized in theoretical frameworks (e.g., Terviö 2009; Waldman 1984).

In this paper I address this missing data challenge directly by building a new dataset that combines the employment histories of newly-minted Ph.D.'s in computer science (CS) with information on their publications in major conference proceedings and their patents. I use the data to show descriptively how the publication of a new paper or a patent application affects inter-firm mobility. I then estimate a structural model of imperfect competition for talent among employers, and use the model to assess the impacts of both public and private learning on the efficiency of talent allocation.

Every year about 4,000 Ph.D.'s graduate in CS or closely related fields in the United States.[1] The majority of new CS Ph.D.'s enter the private sector, but they often continue to publish at academic conferences, yielding public information

---

[1]The number is based on the Survey of Earned Doctorates by the National Science Foundation. Throughout this paper I refer to computer scientists as workers who have a Ph.D. in Computer Science or Electrical Engineering (including EECS) in the United States.

on their research ability.[2] About 25% of papers from industry researchers are accompanied by a patent application filed by their employers: these papers are more highly cited in later years, suggesting that they contain more valuable ideas. The existence of an accompanying patent, however, is private information that is only revealed with an 18-month lag.[3] Patterns of mobility in the period immediately after the patent application (when the fact of filing is private) and in the following few years (when the patent application becomes public information) therefore provide novel evidence of asymmetric learning.

My empirical analysis is based on a new dynamic model of employer learning and sorting that introduces information frictions into a monopsony framework as in Card, Cardoso, Heining, and Kline (2018). I consider the wage setting and task allocation decisions made by forward-looking firms in an imperfectly competitive labor market. Firms that vary in productivity allocate workers to publication-oriented research tasks and update their beliefs about the research ability of workers based on their outputs. When part of the research output is publicly visible, firms face a dynamic trade-off: allocating a worker to publication-oriented tasks allows the firm to benefit from publishing at conferences, but it also increases the risk that high-ability workers will be recognized and poached by outside employers. For simplicity, I assume that workers are myopic and only care about current period wages, though I allow them to have idiosyncratic preferences over different employers. With dynamic incentives solely on the firms' side, I solve for a Markov Perfect Bayesian Nash equilibrium and derive predictions on how employer learning changes the reallocation of workers between firms.

This model generates two key testable predictions: (1) Workers with newly

---

[2]The share of new CS Ph.D.s entering the industry as opposed to academia has been increasing over the past 20 years and exceeding 50% since 2017 (Appendix Figure B3).

[3]The American Inventors Protection Act (AIPA) of 1999 amends title 35, United States Code (U.S.C.) 122 to provide that patent applications shall be published promptly after the expiration of 18 months from the earliest filing date. The United States Patent and Trademark Office (USPTO) has implemented this rule since November 29, 2000.

revealed signals about their productivity in innovation are more likely to move between firms and move to more productive firms than similar workers without such signals. (2) Job mobility is suppressed for workers with positive signals that are observed by the incumbent employer but unknown to potential outside employers. I adopt these predictions as tests for the presence of symmetric (public) and asymmetric employer learning.

The labor market for computer scientists provides rich information on worker productivity that allows me to directly test for employer learning. I match the public LinkedIn profiles of 40,000 computer scientists with their on-the-job research outputs including CS conference proceedings and patent applications. Relative to economics, initial information from the PhD education is less predictive of a computer scientist's future research success.[4] The stronger role of post-PhD employer learning than of initial information in the allocation of talent is also confirmed by a Shapley-value-based decomposition in my structural analysis.

I test for public employer learning by comparing the job mobility of workers who produce a paper with similar coworkers without a paper. I measure upward mobility by job movements into top "big tech" firms {Google, Microsoft, IBM, Facebook, Amazon, Apple} from other nontop firms in the industry.[5] Figure 1a presents a simple comparison between newly minted CS researchers who start off at a nontop firm and either publish or do not publish a paper at a CS conference in the first two years post Ph.D. The raw data clearly shows a divergence in upward mobility rates. Conditional on firm-year fixed effects and a rich set of controls for worker and position characteristics, I find that employees at nontop firms with a paper are more than twice as likely to move to a top firm the next year, suggesting that the revelation of a publication boosts positive assortative matching between

---

[4]I run regressions of post-PhD research accomplishments on PhD school and cohort fixed effects. Using the data on economists in Sarsons (2017), I find a much higher R^2 among economists than among computer scientists (Appendix Table B1).

[5]The top firms pay higher wages and on average produce more papers. About a quarter of CS papers from the industry have an author from the 6 top firms.

3

higher-ability researchers and more productive firms.

Figure 1: Upward Mobility from Nontop to Top Firms

(a) Paper vs. No Paper

(b) Paper+Patent vs. Paper Only
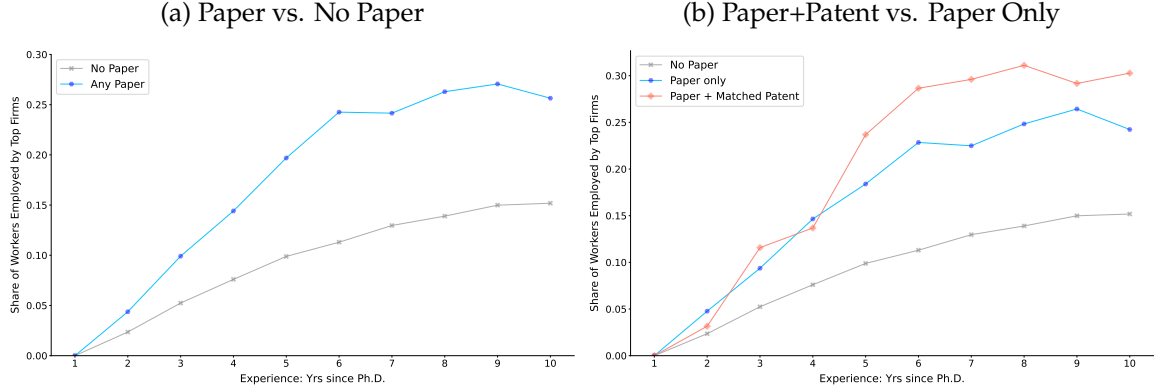
Notes: This figure shows the share of computer scientists who work at a top firm in each year post PhD, separately by a person's research output while working at a nontop firm initially.

To test for asymmetric learning, I exploit patent laws that, by default, delay the disclosure of a patent application by 18 months after its initial filing.[6] This institutional feature suggests whether a paper has a matched patent application is revealed later than the paper itself.[7] Non-disclosure agreements also would not allow workers openly announce pending patent applications that have not been published by the patents office.[8]

Figure 1b shows another divergence in upward mobility between workers who either produce only a paper or a paper with a matched patent. Comparing similar coworkers at nontop firms in an event-study framework, I find that authors of papers with a matched patent are less likely to move than other authors within a year, when only the existence of the paper is public information. But in three years when most patent applications become public information, they are 14%

---

[6]See Title 35 U.S.C. 122 (AIPA 1999) in Appendix Table B4. Figure B1 shows that about 80% of patent applications comply with the 18-month rule. The 20% non-compliance is driven by firms that file a non-publication request at the time of initial filing (see exception B of 122(b) in Table B4).

[7]See Table 2 for examples. I matched patent applications to papers according to the team of authors, employment information, and patentability conditions (Title 35 U.S.C. 102).

[8]Non-disclosure agreements define any invention on the job as the employer's proprietary information. Patent applications that have not been published may even be viewed as trade secrets (*Hyde Corporation v. Huffines* 1958).

more likely to move to a new firm, and 17% more likely to move to a top firm. This finding is consistent with my model of imperfect competition for research scientists: incumbent firms with knowledge of the matched patent would post a higher wage for such workers and therefore retain them longer, but once the matched patent is revealed, public employer learning pulls high-ability workers out of less productive firms. I find similar evidence of employer learning from other mobility outcomes such as moving to a higher-wage firm or promotions (Pastorino 2023). The lack of immediate promotions for authors who have a paper with a matched patent supports the idea that firms may under-place talent to reduce poaching (Milgrom and Oster 1987; Waldman 1984).

How much does employer learning matter for the efficient allocation of labor? To provide a quantitative assessment, I present counterfactual simulations from a fully specified model with and without employer learning from workers' on-the-job research. I estimate the model using a nested fixed-point algorithm as in Rust (1987) to maximize the joint likelihood of job movements and research production by early-career computer scientists. Simulating the model with no learning from papers or patent applications, I estimate that the overall publication rate of CS researchers in the first five years of their career would be 16% lower.

Removing the delayed disclosure of patent applications is estimated to improve publication rate by 1%, which is fully driven by a faster rate of positive assortative matching. Workers who produce a paper with a matched patent would experience a 2 pp increase in upward mobility within a year, and generate a 5-6% increase in innovation production at top firms. However, in the absence of private information rent, incumbent firms would assign fewer publication-oriented tasks ex ante, providing a counterforce on the discovery of talent in this counterfactual scenario. The inefficiency in task allocation is closely related to the prediction that employer-provided general skill training is inefficiently lower when firms have less monospony power (Acemoglu and Pischke 1998; Manning 2003; Stevens 1994).

This paper makes two main contributions. First, I contribute to the employer learning literature by providing direct evidence of the impacts of public learning following publications by CS researchers. Early works by Altonji and Pierret (2001) and Farber and Gibbons (1996) attributed the increasing correlation between wages and AFQT scores (observed by researchers but not firms) over time to employer learning. The underlying model of these studies posits that employers update their belief when new signals arrive, but these signals are rarely observable except from within-firm personnel records (Kahn and Lange 2014). This paper offers more direct tests for public learning by estimating changes in job mobility around a CS publication. Very few articles in this literature test for asymmetric employer learning (Kahn 2013; Schönberg 2007). This paper exploits the delayed disclosure of patent applications to show that workers who produce higher-quality research experience a delayed increase in mobility. Consistent with Hager, Schwarz, and Waldinger (2023), I find that high-ability workers hidden in less productive firms would benefit from a reduction of asymmetric information.

Second, this paper attempts to bring together the theory of employer learning and models of imperfect labor market competition. The classic learning framework often begins with homogeneous players (employers) under perfect competition, which are reasonable simplifying assumptions to discuss complicated problems such as adverse selection (Boozer 1994; Hendricks and Porter 1988; Li 2013). Relaxing the homogeneity and perfect competition assumptions generates a richer set of predictions on job mobility upon information revelation, which I validate in the CS labor market. Doing so does not change the important insight that movers are adversely selected under asymmetric information (Gibbons and Katz 1991; Greenwald 1986). Furthermore, introducing information frictions into a monopsony framework as in Card et al. (2018) provides a tractable model that can be estimated to assess the role of employer learning in the efficient allocation of labor.

2   A Dynamic Model of Employer Learning

I develop a discrete-time finite horizon dynamic model of employer learning by firms in an imperfectly competitive labor market. I first lay out the key assumptions in the conceptual framework, and then fully specify the model and characterize its equilibrium.

2.1   Conceptual Framework

The model concerns the allocation of labor between and within firms given noisy information about workers' binary research ability, denoted by α.

Firms collect revenue from routine activity and from innovation outputs of workers, which may take the form of publications, patent applications, or both, as shown in Figure 2. High-ability workers are more likely than low-ability workers to produce each type of innovation. The likelihood of publishing a paper is jointly determined by a worker's ability α and the share of time she can spend on publication-oriented tasks, τ, which is chosen endogenously by her employer. I allow firms to vary in their returns to different types of innovation. Some firms are better at expanding business through publishing at conferences, whereas other firms may benefit more from private research that yields traditional patents unrelated to papers. The production of publications is supermodular in equilibrium: firms that gain more from publications will set a higher τ, which increases the difference in publication rates between high-ability and low-ability workers.

The challenge of assigning tasks efficiently arises from the uncertainty about worker ability. When new Ph.D.s enter the labor market ($t = 1$), there is public information I_{i1} about person i that is predictive of α_i, such as the prestige of her Ph.D. institution. Post-Ph.D. employer learning, in contrast, is based on the innovation outputs of workers and is asymmetric between firms when the incumbent employer has additional information earlier than the outside labor market. Specifi-

Figure 2: Worker Output
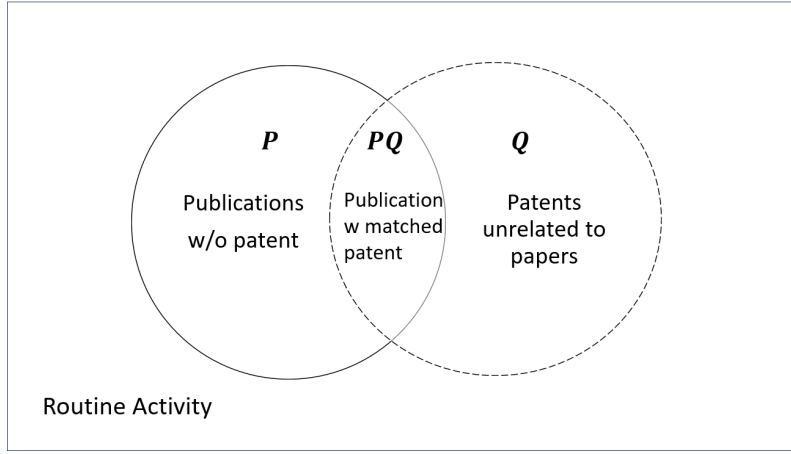
cally, any research publication (solid circle in Figure 2) becomes public information with little or no lag. But whether a paper is accompanied by a patent application, an indicator for higher-quality research, is private information with the employer of the author(s) for one period.

Conditional on information about workers, firms make simultaneous offers of a wage $\{w_{itj}\}$ and a task allocation $\{\tau_{itj}\}$ that maximize expected flow profit plus a discounted continuation value from the worker. Importantly, they face a dynamic tradeoff: setting a higher τ can increase a firm's revenue from publications today, but it also increases the risk that other firms recognize and poach the authors in the next period. Such turnover risk is higher at less productive firms, which post lower wages on average in equilibrium. The downward pressure of turnover risk on publication-oriented tasks is the same as how monopsony power affects employer-provided general skill training (Acemoglu and Pischke 1998; Manning 2003; Stevens 1994).

To focus on the dynamic decisions by firms, I keep workers' problem simple and static. At $t = 1$, workers observe the wage postings and draw idiosyncratic preferences over employers, ϵ_{i1j}, which can be correlated within each nest of employers $G(j) \in \{$Tenure-Track, Postdoc, Top Firms, Nontop Firms$\}$ but independent

8

between G's. At $t > 1$, I follow Card et al. (2018) to let workers re-enter the job market and redraw preferences with probability $\lambda(I_{it})$, which is a function of the public information I_{it} known to all employers at the beginning of t. Other workers are assumed to stay with their original employers. When $\lambda < 1$, firms have additional monopsony power over their incumbent employees and can set lower wages than for equally productive new workers.

I show the existence and uniqueness of a Markov Perfect Bayesian Nash Equilibrium in the dynamic contract-posting game between firms. The wage increase upon public information revelation is higher at more productive firms, pushing high-ability workers up along the firm job ladder. The simplifying assumption that workers naively solve a static job choice problem shuts down self-selection into more research-intensive jobs (Stern 2004), but they do not change the key model predictions (Section 2.3) on increased mobility from less productive firms when the labor market receives positive information about workers.

2.2 Model Specification and Equilibrium

I clarify the notation and the information structure, state the repeated static problem of workers, the dynamic problem of firms, and solve for the equilibrium in this finite T-period game via backward induction.

2.2.1 Notation and Information Structure

Production. Denote by (P, PQ, Q) the three types of innovation in Figure 2: P represents publications without a matched patent, PQ represents publications with a matched patent, and Q represents patents unrelated to research publications. The innovation output of a worker per period is summarized by a vector $(D_{it}(P), D_{it}(PQ), D_{it}(Q))$, each of which indicates if i produces that type of innovation during period t. Per unit of time on publication-oriented tasks, high-ability workers produce a paper with probability p_H, which is assumed to be higher than

9

p_L, the probability for low-ability workers. The publications produced by high-ability workers are also more likely to be high-quality and have a matched patent application, $p_H^* > p_L^*$. As shown in Table 1, the likelihood of any research publication is jointly determined by worker ability and publication-oriented tasks τ, and it does not depend on the identity of employer j conditional on τ. High-ability workers are also more capable of producing patents unrelated to papers, but the likelihood of producing Q is independent from τ. For simplicity, I do not let firms endogenously assign a patent-oriented task but allow firms to vary in baseline patenting rates denoted by $\bar{q}_j$.[9]

Table 1: Likelihood of Innovation Output

	Innovation Output	Likelihood
$D_{it}(P)$	Any Paper but no Matched Patent	$E[D_{it}(P)\,|\alpha,\tau,j] = p_\alpha \times (1 - p_\alpha^*) \times \tau$
$D_{it}(PQ)$	Any Paper + Matched Patent	$E[D_{it}(PQ)|\alpha,\tau,j] = p_\alpha \times p_\alpha^* \times \tau$
$D_{it}(Q)$	Any Patent unrel. to Paper	$E[D_{it}(Q)\,|\alpha,\tau,j] = q_\alpha + \bar{q}_j$

Firms are endowed with a baseline productivity $\bar{\phi}_j \in \mathbb{R}^+$, and proportionate returns to each type of innovation, $\left[\phi_j(k)\right]_{k \in \{P, PQ, Q\}}$ with $\phi_j(k) \in \mathbb{R}^+$, all of which are publicly known. Firms that benefit more from publications have a higher $\phi_j(P)$ or $\phi_j(PQ)$, while firms that rely more on patenting have a higher $\phi_j(Q)$. The expected value from the production at firm j conditional on worker ability α and

---

[9]$\bar{q}_j$ is a normalization of the firm's return to patents unrelated to papers, $\phi_j(Q)$. I calibrate $\phi_j(Q)$ based on firm fixed effect in patenting before estimating the model (Section 5).

the share of time allocated to publication-oriented tasks, τ, is:[10]

$$Y_j(\alpha, \tau) = \bar{\phi}_j \left(\underbrace{1 - \tau}_{\text{routine}} + \underbrace{\sum_{k \in \{P, PQ, Q\}} \phi_j(k) \times E[D_{it}(k)|\alpha, \tau, j]}_{\text{returns to innovation}} - \underbrace{\zeta(\tau)}_{\text{cost}} \right) \qquad (2.1)$$

Information Structure. The payoff-relevant state space for firms is defined by the information about workers. Denote by I_{it} the public information about the research ability of worker i at the beginning of t, and by $\widetilde{I}_{it}$ the private information known only to her incumbent employer. At $t = 1$, I_{i1} includes her education and publication records before Ph.D., and $\widetilde{I}_{i1} = \emptyset$. Once a worker has entered the labor market, information evolves according to her on-the-job innovation output.

Employer $j(i, t)$ of worker i in period t will have full access to her innovation output $(D_{it}(P), D_{it}(PQ), D_{it}(Q))$ in that period. However, by the beginning of the next period, the outside employers will only know if there is a paper published during t, denoted by $D_{it}(P) + D_{it}(PQ)$, but cannot tell if her publication has a matched patent application, or if she has other patents unrelated to papers (see the dashed circle in Figure 2). Under the assumption that $\widetilde{I}_{it}$ becomes public with a

---

[10]There is a convex cost of allocating workers to publication-oriented tasks, which may include investment in computing power that often grows in a convex way as employees spend more time on research. It may also absorb the management costs of moving workers away from routine activities within a firm. For example, a firm may have to establish an in-house research lab, host academic consultants, and establish a new performance evaluation system for workers who are increasingly involved in research.

one-period delay, the information evolution is summarized as follows:[11]

$$\text{public } I_{i(t+1)} = \underbrace{I_{it} \cup \widetilde{I}_{it}}_{\text{info before } t} \cup \{j(i,t), \underbrace{D_{it}(P) + D_{it}(PQ)}_{\text{any paper at } t}\} \tag{2.3}$$

$$\text{private } \widetilde{I}_{i(t+1)} = \{(D_{it}(P), D_{it}(PQ), D_{it}(Q))\}$$

The model timeline is detailed in Appendix A0. At least three discrete periods are needed to capture the full information revelation process.

2.2.2 Workers' Problem

Workers who are on the labor market at t draw idiosyncratic preferences from a generalized extreme value (GEV) distribution:

$$F(\{\epsilon_{itj}\}) = exp\left(-\sum_{G \in C}\left(\sum_{j \in G} exp(-\rho_G^{-1}\epsilon_{itj})\right)^{\rho_G}\right), \rho_G \in (0,1] \tag{2.4}$$

where C denotes the set of potential employers a worker can choose from in a given period.[12] Under this assumption, preferences are independent between nests and over time, but can be correlated within a nest G if $\rho_G < 1$. Among the four nests $G(j) \in \{\text{Tenure-Track, Postdoc, Top Firms, Non-Top Firms}\}$ in the CS labor market, the first two represent academia while the last two represent industry.

All workers are on the labor market at $t = 1$ (the first year post PhD). At $t > 1$ any worker i from nest G with public information I_{it} can get on the market again and search for new jobs with probability:

$$\lambda(I_{it}) = \lambda_{0,G} \times (1 + \lambda_{1,G} \times Pr(H \mid I_{it})) \tag{2.5}$$

---

[11]The conditional probability distribution of future states depends only on the current state, satisfying the Markov property:

$$Pr(I_{i(t+1)}, \widetilde{I}_{i(t+1)}|I_{it}, \widetilde{I}_{it}) = \sum_{\alpha \in \{H,L\}} \underbrace{Pr(\alpha|I_{it}, \widetilde{I}_{it})}_{\text{current belief}} \times \underbrace{Pr(D_{it}(P), D_{it}(PQ), D_{it}(Q)|j(i,t), \alpha)}_{\text{innovation output at } t} \tag{2.2}$$

[12]Workers from industry may not be always be able to move to academia. In that case, C does not include tenure-track or postdoc employers. See footnote [58].

which takes a positive value in $(0, 1]$, and can vary between original nest G's and depend on public belief $Pr(\alpha_i = H \mid I_{it})$ about the worker.[13][14] Other workers who are not on the market stay put and hold fixed the preferences they have drawn before.

Workers who are on the labor market observe the wages posted simultaneously by potential employers $\{w_{itj}\}$ and choose an employer as follows:

$$j(i, t) = argmax_{j \in C} \ u_{itj} = b \times ln(w_{itj}) + \rho_{G(j)} \times \epsilon_{itj} \tag{2.6}$$

Assume b is positive and finite and $\forall G : \rho_G \in (0, 1]$. The elasticity of labor supply increases in the ratio $\frac{b}{\rho_G}$. Throughout the analysis below, I assume this ratio is finite, which means the the labor market is imperfectly competitive. Given GEV preference shocks, a worker's choice probabilities are represented by the well-known nested logit model (McFadden 1973; Imbens and Wooldridge 2007):

$$s_{j|C} = \underbrace{s_{j|G(j)}}_{\text{choose } j \in G(j)} \times \underbrace{s_{G(j)|C}}_{\text{choose nest } G(j) \in C} \tag{2.7}$$

each of which is a function of wages within a choice set C. Conditional on public information I_{it} and posted wages $\{w_{itj}\}$, the worker's expected labor supply to her incumbent employer vs. to an outside employer can be written as:

Incumbent $j = j(i, t-1)$: $\ s_j^{(1)}(\{w_{itj'}\}; I_{it}) = \underbrace{1 - \lambda(I_{it})}_{\text{off market}} + \underbrace{\lambda(I_{it}) \times E_C[s_{j|C}]}_{\text{on market \& choose j again}}$

Outside $j \neq j(i, t-1)$: $\ s_j^{(0)}(\{w_{itj'}\}; I_{it}) = \lambda(I_{it}) \times E_C \left[s_{j|C} \right] \tag{2.8}$

The elasticity of labor supply to a firm is lower among incumbent employees when

[13]For example, a worker with higher market belief but employed by a low-productivity firm may search for new jobs more frequently, in which case $\lambda_{1,G} > 0$ for $G = $ Non-Top Firms. Workers from top firms, in contrast, may be less likely to search for new jobs when they are perceived as high-ability by the market.

[14]This formulation is equivalent to each worker drawing a random search cost $z \overset{d}{\sim} \Phi$, and only search for new jobs if $z < \bar{z}$, where $\Phi(\bar{z}) = \lambda$. The λ's can also be interpreted as job arrival rates in search models (e.g. Burdett and Mortensen 1998;Postel-Vinay and Robin 2002).

$\lambda < 1.$[15] The labor market frictions arising from the fact that workers are not always on the market gives employers additional monopsony power over incumbent employees relative to new workers.

2.2.3 Employers' Problem

I focus on how employers set wages and allocate workers to publication-oriented tasks in an intermediate period $t \in \{2, , ..., T-1\}$. The complete backward induction is presented in the Appendix A1. Employers see either public information (I_{it}) about workers from other firms, or public and private information $(I_{it}, \widetilde{I}_{it})$ about incumbent workers. They do not know whether a worker is on the market or not, or her specific preferences, and therefore cannot price discriminate accordingly.

For an incumbent employee with public and private information $(I_{it}, \widetilde{I}_{it})$, employer j solves for an optimal contract $(w_{itj}^{(1)}, \tau_{itj}^{(1)})$, taking as given the wages set by other firms, denoted by w_{-j}:[16]

$$v_{tj}^{(1)}(I_{it}, \widetilde{I}_{it}) = max_{w,\tau} \underbrace{s_j^{(1)}(w, w_{-j}; I_{it})}_{\text{expected labor supply}} \times \underbrace{\left(E_{\alpha|I_{it}\cup\widetilde{I}_{it}}[Y_j(\alpha, \tau)] + \beta E[v_{(t+1)j}^{(1)}(I', \widetilde{I}') | \tau] - w\right)}_{\text{flow profit \& discounted continuation value}}$$

(2.10)

The continuation value equals the value from an incumbent worker $v_{(t+1)j}^{(1)}$, expected

---

[15]The elasticity of an incumbent worker vs. an outside worker to firm j, given public information I and the wages posted by other firms:

$$\xi_{itj}^{(1)} := \frac{\partial ln(s_j^{(1)}(w, w_{-j}; I))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C[\frac{\lambda(I) \times s_{j|G} \times s_{G|C}}{s_j^{(1)}} \times (1 - \rho_G s_{j|G} s_{G|C} - (1 - \rho_G)s_{j|G})] \quad (2.9)$$

$$\xi_{itj}^{(0)}(\widetilde{I}) := \frac{\partial ln(s_j^{(0)}(w, w_{-j}; I, \widetilde{I}))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C\left[\frac{s_{G|C}}{E_C[s_{G|C}]} \times (1 - \rho_G s_{j|G} s_{G|C} - (1 - \rho_G)s_{j|G})\right]$$

Note $\lambda(I)$ does not show up in the elasticity of an outside worker.

[16]In equilibrium (Definition 1, the wage $w_{tj}^{(1)}(I, \widetilde{I})$ for an incumbent employee is the best response to $w_{-j} = w^{-j}(I)$ conditional on public information I.

14

over the innovation outputs that will be produced in the current period.[17] It is discounted by a common factor β.

The optimal wage in this dynamic problem is front-loaded with the expected continuation value from a job stayer, and is marked down by the inverse of labor supply elasticity $\xi_{itj}^{(1)}$ (2.9):[18]

$$w_{itj}^{(1)} = \left(E_{\alpha|I_{it} \cup \widetilde{I}_{it}}[Y_j(\alpha, \tau)] + \beta\, E[v_{(t+1)j}^{(1)}(I', \widetilde{I}') \mid \tau] \right) \times \underbrace{\xi_{itj}^{(1)} \times \left(1 + \xi_{itj}^{(1)} \right)^{-1}}_{\text{markdown}} \qquad (2.11)$$

Publication-oriented task allocation is chosen to maximize the expected returns to publications today and dynamic returns to the continuation value. Employers consider how task allocations would affect public information about a worker and her turnover tomorrow.

$$\tau_{itj}^{(1)} = max\{0, min\{1, \tau_{tj}^*(I_{it}, \widetilde{I}_{it})\}\} \qquad (2.12)$$

$$\tau_{tj}^*(I, \widetilde{I}) := \frac{1}{\zeta} \times \underbrace{E_{\alpha|I \cup \widetilde{I}}\left[-1 + \sum_{k \in \{P, PQ, Q\}} \phi_j(k) \times \frac{\partial E[D_{it}(k)|\alpha, \tau]}{\partial \tau} \right]}_{\text{return to innovation today}} + \frac{\beta/\bar{\phi}_j}{\zeta} \times \underbrace{\frac{\partial E[v_{(t+1)j}^{(1)}(I', \widetilde{I}')|\tau]}{\partial \tau}}_{\text{dynamic return}}$$

The derivative of continuation value over task allocation can be negative if workers who successfully publish are likely to leave the original firm.[19] The lower assignment of publication-oriented tasks in that case is similar to the inefficiently lower training provided by firms that face higher turnover (e.g. Acemoglu and Pischke 1998; Stevens 1994).

For an outside worker i from $j(i, t-1) \neq j$, firm j only has access to public

---

[17]Information from t to $t+1$ evolves according to the worker's innovation outputs at t as in (2.3), with $I' = I_{it} \cup \widetilde{I}_{it} \cup \{D_{it}(P) + D_{it}(PQ)\}$, $\widetilde{I}' = \{(D_{it}(k))_{k=P,PQ,Q}\}$. See the continuation value in (7.16).

[18]Incumbent employers can set a higher wage for workers who are privately known to be better than what the market believes. The higher wage itself (posted simultaneously) would not disclose private information directly.

[19]The dynamic return, as expressed in (7.20), is negative when the cost of losing a worker exceeds the benefits to internal reallocation of talent in the next period.

information I_{it}. Its value function is expected over the unknown private signals $\widetilde{I}_{it}$:

$$v_{tj}^{(0)}(I_{it}) = max_{w,\tau} E_{\widetilde{I}|I_{it}} [\underbrace{s_j^{(0)}(w, w_{-j}; I_{it})}_{\text{expected labor supply}} \times \underbrace{\left(E_{\alpha|I_{it}\cup\widetilde{I}}[Y_j(\alpha, \tau)|I_{it} \cup \widetilde{I}] + \beta E[v_{(t+1)j}^{(1)}(I', \widetilde{I}')|\tau] - w\right)}_{\text{MRPL \& discounted continuation value, net wage}}]$$

(2.13)

The wage for a new worker is a weighted average of monopsonistic wages (marked down by elasticity $\xi_{itj}^{(0)}$ (2.9) conditional on information (I_{it}, $\widetilde{I}$):

$$w_{itj}^{(0)} = \left(1 + E_{\widetilde{I}|I_{it}}\left[\frac{s_j^{(0)}}{E_{\widetilde{I}|I_{it}}[s_j^{(0)}]} \times \xi_{itj}^{(0)}(\widetilde{I})\right]\right)^{-1}$$

(2.14)

$$\times E_{\widetilde{I}|I_{it}}\left[\frac{s_j^{(0)}}{E_{\widetilde{I}|I_{it}}[s_j^{(0)}]} \times \xi_{itj}^{(0)}(\widetilde{I}) \times \left(E_{\alpha|I_{it}\cup\widetilde{I}}[Y_j(\alpha, \tau)] + \beta E[v_{(t+1)j}^{(1)}(I', \widetilde{I}')|\tau]\right)\right]$$

in which the weight on each possible $\widetilde{I}$ equals to the probability of $\widetilde{I}$ being the private information given public I_{it} and the fact that the worker moves into j.[20] When incumbent employers set higher wages for workers with more positive $\widetilde{I}$, outside firms would lower the weights on such $\widetilde{I}$, taking into account that movers under asymmetric information are adversely selected (Gibbons and Katz 1991; Greenwald 1986). Such weight adjustments lower the wages posted by outside employers, at least partially correcting for the winner's curse (Boozer 1994; Li 2013; Hendricks and Porter 1988).

The optimal allocation of new workers to publication-oriented tasks is a weighted average of $\tau_{tj}^*(I, \widetilde{I})$, with the same weight on each unknown $\widetilde{I}$ as in (2.14):

$$\tau_{itj}^{(0)} = E_{\widetilde{I}|I_{it}}\left[\frac{s_j^{(0)}}{E_{\widetilde{I}|I_{it}}[s_j^{(0)}]} \times \tau_{tj}^*(I_{it}, \widetilde{I})\right], \quad \tau^* \text{ defined in (2.12)}$$

(2.15)

---

[20]Since the wages enter the weights on the right-hand side, there are no analytic solutions, but I will show the equilibrium wages can be solved via fixed-point iterations.

2.2.4 Equilibrium

I define a Markov Perfect Bayesian Nash Equilibrium (MPBNE) in this finite-horizon discrete-time game, in which firms post profit-maximizing contracts conditional on their current information about a worker, taking as given the contracts posted by other firms.

Definition 1 (Markov Perfect Bayesian Equilibrium Under Imperfect Competition)
Given that workers are impatient and choose employers based on wages and GEV-distributed preferences (2.6), a strategy profile $\{(w_{tj}, \tau_{tj}) : t = 1...T\}$ is a Markov Perfect Bayesian Nash Equilibrium in the finite-horizon game between firms if it satisfies:

- *$(w_{tj}^{(1)}(I, \widetilde{I}), \tau_{tj}^{(1)}(I, \widetilde{I}))$ maximize the expected value from an incumbent employee given any information $(I, \widetilde{I})$ at the same period;*

- *$(w_{tj}^{(0)}(I), \tau_{tj}^{(0)}(I))$ maximize the expected value from an outside worker given any public information I and rational belief about private information $\widetilde{I}$ conditional on I.*

I can write the equilibrium in Definition 1 as a system of equations on the allocation of workers between firms, which are functions of the posted wages.[21] Let w denote the vector of wage strategies in equilibrium, and $s = s(w)$ denote the vector of expected labor supply (2.8) evaluated at w. The equilibrium can be defined by:

$$s = s(w(s)) \tag{2.16}$$

in which s, the vector of information-state-specific probability of workers choosing each firm, is a fixed point of the composite function $s \circ w$. Following Rust (1994), I show $s \circ w$ is a contraction mapping and therefore the MPBNE exists. I further show the uniqueness of the choice probabilities (or wages up to scaling by positive constants) given that workers make a static choice between firms based on wages and idiosyncratic preferences (Proposition 1 in Appendix A2).

---

[21]The choice probabilities in equilibrium are not functions of publication-oriented tasks since workers only value wages and idiosyncratic preferences. We relax this restriction in the extension with forward-looking workers (Appendix A4).

The allocation to publication-oriented tasks in equilibrium is very similar to firms' provision of general skill training. Firms assign more ability-revealing tasks when information about workers is less public (Acemoglu and Pischke 1998), and when they have more monopsony power (Manning 2003; Stevens 1994). Under perfect competition ($\frac{b}{\rho_G} \rightarrow \infty$), workers who are not credit-constrained bear all costs of ability-revealing tasks and are paid their full marginal product of labor as in Becker (1964) (see Proposition 2 in Appendix A2).

2.3   Model Predictions

I discuss the implications of information revelation on inter-firm mobility. I focus on workers from less productive firms, and make an additional assumption: in the nest G = nontop firms, the probability of re-entering the labor market (2.5) satisfies: $\forall$ information I, I': $Pr(\alpha = H|I) > Pr(\alpha = H|I') \rightarrow \lambda(I) > \lambda(I')$. Workers from nontop firms are at least as likely to re-enter the job market as coworkers with a lower market belief about their ability. In the predictions below, the changes in job mobility when new information becomes public are driven by the changes in the wages posted by employers, rather than arbitrary changes in λ's at which workers get on the job market.

Prediction 1 (Job Mobility in Response to Publications) *Conditional on public information about research ability, workers who publish a paper while being employed by less productive firms are*

  a) more likely to move to a new firm,
  b) more likely to move to firms with higher returns to innovation

in the next year than coworkers without a publication.

Publications improve the market belief about if a worker is H-ability. The equilibrium wages across firms increase in response to the positive signal. But importantly, firms with a higher return to publications $\phi_j(P\cdot)$ assign more publication-oriented tasks to the same worker. The complementarity between firms and workers

in equilibrium leads to a larger wage increase at more productive firms. As a result, the model predicts an increase in inter-firm mobility, and an increase in upward mobility for workers who publish a paper at less productive firms.

Prediction 2 (Job Mobility under Asymmetric Information: $D_{it}(P)$ vs. $D_{it}(PQ)$)
Workers who have produced a high-quality paper with a matched patent, i.e. $D_{it}(PQ) = 1$, while being employed by less productive firms are

  a) *less likely to leave employer $j(i, t)$ when the presence of a matched patent $D_{it}(PQ) = 1$ is private information;*

  b) *more likely to leave and move upward after $D_{it}(PQ) = 1$ is revealed.*

than coworkers with papers but no matched patents $D_{it}(P) = 1$.

The second prediction relies on the assumption that at the beginning of $(t + 1)$, the outside labor market can see any publication during t but cannot differentiate between $D_{it}(P)$ and $D_{it}(PQ)$. Incumbent employers set a higher wage based on a more favorable private belief $Pr(H|I \cup \widetilde{I}) > Pr(H|I)$ when $D_{it}(PQ) = 1$, and therefore reduces the turnover of $D_{it}(PQ) = 1$ workers relative to $D_{it}(P) = 1$ coworkers who produce papers without a matched patent. Once the matched patent is revealed, Prediction 1 applies (see the complete proof in Appendix A2).

Prediction 3 (Job Mobility under Asymmetric Information: $D_{it}(Q)$) *Workers who produce a patent application unrelated to any paper, $D_{it}(Q) = 1$, experience a delayed increase in job mobility relative to similar coworkers with $D_{it}(Q) = 0$.*

This prediction can be viewed as a corollary of Prediction 2. Patents unrelated to papers, denoted by Q, are also revealed with a delay. $D_{it}(Q) = 1$ workers are also predicted to stay at the incumbent longer but move once the positive signal becomes public information.

I test the model predictions in Section 4, and estimate the model to quantify the role of employer learning in the efficiency of talent (re)allocation in Section 5.

3   Data

I collected data on the career trajectories and research outputs of Ph.D. computer scientists. This section discusses the matching between Ph.D. dissertation records and public LinkedIn profiles, and measures of on-the-job research that include conference papers and patent applications.

3.1   Ph.D. Graduates in Computer Science

I focus on Ph.D. graduates in CS or closely related fields, who, like economics Ph.D.s, may take a tenure-track or postdoc job in academia, or a job outside academia that can also be research-intensive.[22] The share of new CS Ph.D.s entering the industry as opposed to academia has been increasing over the past 20 years and has been over 50% since 2017 (Appendix Figure B3). On the ProQuest Theses and Dissertation Database, I found about 81,000 Ph.D. dissertations in Computer Science or Electrical Engineering from the top 60 CS schools in the United States, between 1980 and 2021.[23] Each dissertation provides the full name of the doctoral recipient, school, and year of PhD.[24]

3.2   Public LinkedIn Profiles of CS Ph.D.'s

I developed a program that acts as a recruiter and views public profiles on LinkedIn, the largest online professional network. For each person in the dissertation sample, I submitted a web query that searches by the person's name, PhD

---

[22]See Appendix Figure B2 for research scientist job ads. CS Ph.D.s may also work as engineers, but they often start as senior software engineers or as research engineers who can also publish papers.

[23]I use the U.S. ranking of computer science departments on CSRankings, which is developed and maintained by Emery Berger at UMass Amherst.

[24]Appendix Table B2 displays the number of dissertations by year.

institution, and degree information.[25] About 51% queries returned at least one LinkedIn profile, resulting in 41,000 fully matched profiles in total.[26]

Each profile is formatted as a résumé. The program collects public information such as education background and employment history. I constructed a longitudinal dataset of post-Ph.D. job history for the fully matched LinkedIn profiles. On average a CS Ph.D. has 2.1 industry employers, 0.2 postdoc employers, and 0.3 academic (tenure-track) employers after Ph.D. (Table B5). For each person×year, I kept the primary employer and the corresponding job title.[27] A job-to-job movement in year t is defined as a change in one's primary employer in comparison with her employer next year: $j(i, t) \neq j(i, t + 1)$. Years without any employer would not be considered as a job movement. About 12% of workers at non-top firms move to a new employer per year, whereas workers at top firms or in academia are less mobile (Table B6).

3.3 On-the-job Research

3.3.1 Computer Science Papers

To measure the research output of Ph.D. computer scientists, I collected papers published in 80 CS conferences and two machine learning journals, which are used to rank CS departments (CSRankings). I search for papers at each conference/journal×year on Scopus, a large-scale publication database produced

---

[25]Appendix Figure B5 shows a sample query on LinkedIn Recruiter Platform. A profile is considered matched to the PhD graduate only if the full name and PhD institution are matched exactly, and the year of Ph.D. is the same whenever available on the profile.

[26]See Appendix B for details. The matching rate is higher for more recent cohorts (Figure B3). LinkedIn was first launched in 2003, and its members grew from 37 million in 2009 to 875 million in 2023.(https://www.businessofapps.com/data/linkedin-statistics/).

[27]If there is more than one employer in a year, I rank the jobs in the order of 1) full-time position (over contract or visiting), 2) number of months on the job during the year, and 3) tenure at the employer.

by Elsevier.[28] Each paper comes with a complete list of authors and their affiliations, which indicate the employer of an author at the time of publication.[29]

To match papers with education and job histories, I cleaned and harmonized author affiliations listed on CS papers, Ph.D. institutions and employers on LinkedIn. A paper matched to an author's Ph.D. institution by the time she graduates is labeled as pre-Ph.D. research. After Ph.D., a paper is considered as on-the-job research if the author affiliation matches with her employer at the time of publication.[30] About 28% of matched computer scientists have at least one on-the-job research publication post Ph.D. (Table B5). The publication rate at person-year level is higher at top firms: 10% of employees of top firms publish a paper per year, versus 2% of employees of non-top firms (Table B6).

3.3.2   Paper-Patent Matches

Firms often file patent applications to protect the inventions disclosed in a research publication. I establish a potential paper-patent linkage if the following conditions are satisfied:

1. The majority of authors in the paper are also inventors in the patent application and vice versa.
2. A patent assignee can be matched to an author's affiliation on the paper, which is also her current employer as shown on LinkedIn.
3. The patent application is initially filed between $[-2, 1]$ years relative to the publication of the paper (using conference date).[31]

---

[28]I am especially grateful to Anna Sun for her help with the Scopus Search API.

[29]I cross-validate the data from Scopus by merging it with paper-author records on DBLP, a popular computer science bibliography (see Appendix B for details).

[30]The publication cycle is significantly shorter in computer science. It is unlikely for a dissertation chapter to be published as a conference proceeding years later. I further check if coauthors on a paper are affiliated with the Ph.D. school or with the current employer. Roughly 1% of post-PhD publications have the majority of coauthors affiliated with the Ph.D. school, and are excluded from on-the-job research production.

[31]The patent laws in the U.S. allow the inventors to apply for a domestic patent for inventions that

4. Text is similar: the l^2 norm between the vector embedding of the paper's title plus abstract and the embedding of the patent's is ≤ 0.33. [32]

For each paper, I keep the best matched patent application based on the number of team members in common, similarity between the abstracts, and the closeness of timing. About 25% of papers by CS Ph.D.'s working in the industry, and 5% of papers by those in academia, are accompanied by a patent application. 90% of the matched patent applications are filed before the research paper shows up at a conference. Table 2 and Appendix Table B7 provides examples of paper-patent matches. A matched pair may have different titles, but the abstracts are similar. CS papers tend to be shorter, while patent applications contain more technical details and are more precise about contributions that can be claimed as inventions.

Table 2: Examples of CS Papers and Matched Patent Applications

Firm	Team	Text Distance	Paper	Matched Patent Application
Yahoo	100%	0.233	UNBIASED ONLINE ACTIVE LEARNING IN DATA STREAMS; 08/2011	ONLINE ACTIVE LEARNING IN USER-GENERATED CONTENT STREAMS; Filed: 10/2011, Published: 05/2013
Adobe	80%	0.273	FORECASTING HUMAN DYNAMICS FROM STATIC IMAGES; 07/2017	FORECASTING MULTIPLE POSES BASED ON A GRAPHICAL IMAGE; Filed: 04/2017, Published: 10/2018

Figure 3 shows that papers with a matched patent are higher quality on average. In the first year, they receive roughly the same number of citations as those without a matched patent. But the gap starts to expand around two years

---

are disclosed in any publication no earlier than a year ago. In most other countries, inventions that have been disclosed, for example via a research paper, cannot be filed as a patent application.

[32]The text embedding of a paper or a patent application was obtained via GPT4-ada, a state-of-the-art large language model trained by OpenAI. The threshold for the distance between a paper's embedding and a patent's embedding is selected based on the ROC curve in Appendix Figure B6 to balance between false positive and false negative rates.The norm of an embedding vector is one. Ranking paper-patent pairs by $l2$-norm between vectors is equivalent to ranking them by cosine distance.

Figure 3: Citations Received by Papers With vs. Without a Matched Patent Application

Citation Path: Papers with vs. without a Matched Patent Application
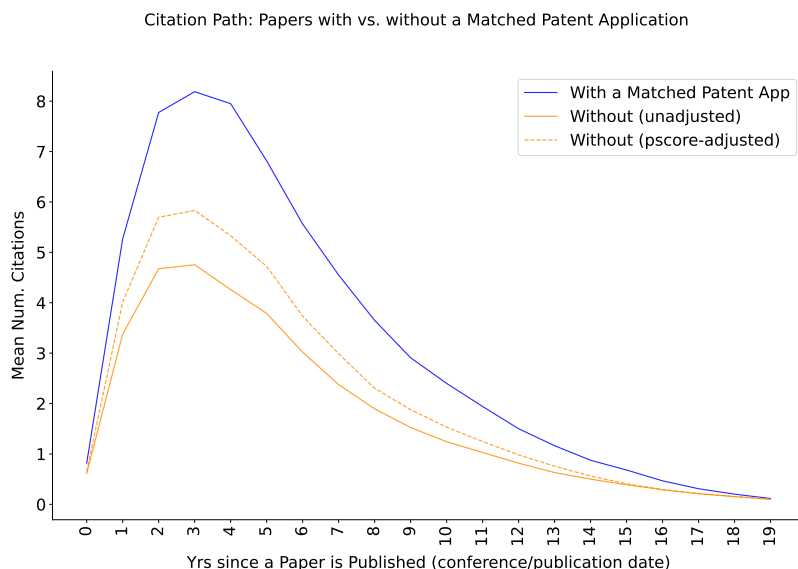
Notes: See Appendix B for details on the measure of paper citations and the reweighting procedure to adjust for firm-year heterogeneity in patenting a CS paper.

after the paper becomes public, which coincides with the disclosure of patent applications (Appendix Figure B1). The quality difference between papers with or without a matched patent suggests: 1) incumbent employers have additional information about the quality of a paper and can act on it by filing for a patent, 2) it takes time for outsiders to recognize valuable research and the divergence in citations happens around the same time as the revelation of a matched patent.

3.3.3 Other Patent Applications

To obtain a more complete picture of innovation activity, I merged the panel of CS Ph.D.s with US patent applications that are not related to papers (Q in Figure 2). I require the assignee (firm) on the application to be the same as the inventor's employer as shown on LinkedIn, and the year of the initial filing to fall within the years she works at the firm. Over 40% of the computer scientists have at least one

patent application after PhD (Table B5).

4 Empirical Tests for Employer Learning

I present reduced form evidence - based on an event study style analysis - for the presence of both public and private employer learning.

4.1 Public Learning: Mobility Responses to CS Papers

Figure 4: Differences in Inter-firm Mobility: With vs. Without a Paper

(a) Any Move between Firms   (b) Employment by Top Firms Next Year
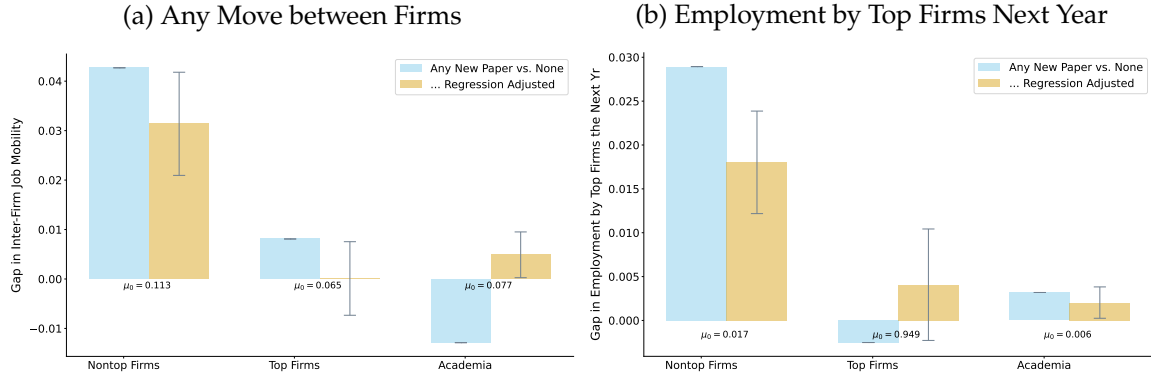
Notes: The blue bars are unadjusted raw gaps in job mobility, whereas the yellow bars are adjusted in a regression that controls for Ph.D. school, experience since Ph.D., position types, and firm-year fixed effects. μ_0 refers to the mean mobility among workers without a new CS paper.

To test for public learning in Prediction 1, I compare the job mobility between workers who produce a CS paper and their coworkers without a paper. Figure 4(a) first shows the raw difference (blue bars) in inter-firm mobility between these two groups. At nontop firms, workers who produce a paper are on average 4 pp or 35% more likely to move to a new firm the next year. This difference remains significant (yellow bar) when I control for firm-year fixed effects to compare coworkers at the same firm in the same year, and additional worker characteristics such as PhD school and cohort, experience and current position types. With the same set of controls, I also find a significant but smaller increase in mobility when workers in

25

academia publish a new paper, but there is no change in inter-firm mobility for workers who are already at the top firms.[33] Given the lower publication rate on average at nontop firms, I interpret them as less innovative firms relative to the top firms. The increase in job mobility when employees at nontop firms publish a paper provides evidence for public employer learning in Prediction 1(a).

Workers who publish at nontop firms are also more likely to move up the job ladder (Figure 1). This pattern is further confirmed by the regression-adjusted estimates in Figure 4(b). Conditional on origin, workers with a publication are twice as likely to move from a nontop firm to a top firm the next year relative to similar coworkers. The results are also robust when conditioning on person fixed effects (Appendix Table C2). Academics who publish are also more likely to move to top firms but at much lower rates. Publications appear to help employees at top firms stay within the top tier, but the difference is not statistically significant.

4.2   Asymmetric Learning: Papers vs. Patent Applications

To provide evidence on the importance of asymmetric learning, I use an event-study framework to analyze job mobility in the years following a new paper/patent combination. Recall that the combinations shown in Figure 2 are denoted by: $D_{it}(P) = 1$ if a worker has CS publications but none of which are matched to patent applications, $D_{it}(PQ) = 1$ if she has any paper with a matched patent application, and $D_{it}(Q) = 1$ if she has a newly filed patent with no related paper.

Due to the delayed disclosure of patent applications, in the index year t the outside employers cannot differentiate between $D_{it}(P)$ and $D_{it}(PQ)$ when they see

---

[33]"Academia" includes postdocs and professors. The raw difference in academia is negative in Figure 4(a), but it is driven by the fact that professors publish papers at higher rates than postdocs but are less mobile. Once I control for position type, I find a 0.5 pp significant increase in job mobility in academia.

a publication. Whether a worker applies for patents unrelated to papers, $D_{it}(Q)$, is also not revealed immediately. In three years, 95% of the patent applications will become public information (Appendix Figure B1). Define Lagged-$D_{it}(PQ) = 1$ if a worker produces any paper with a matched patent application in the past three years. It can be differentiated from Lagged-$D_{it}(P)$, which indicates having a publication but no matched patent during $[t-3, t-1]$. I estimate:

$$M_{it} = \underbrace{\sum_{k \in \{P,PQ,Q\}} \beta_k \times D_{it}(k)}_{\text{signals about new outputs}} + \underbrace{\sum_{k \in \{P,PQ,Q\}} \gamma_k \times \text{Lagged-}D_{it}(k)}_{\text{signals about lagged outputs } [t-3,t-1]} \qquad (4.1)$$

$$+ \underbrace{W_{it}' \Gamma}_{\text{controls}} + \underbrace{\mu_{j(i,t),t}}_{\text{firm-yr}} + \xi_{it}$$

where M_{it} is a mobility outcome at person×year level, which can be any movement between firms, or a movement into a top firm. The firm-year fixed effects, denoted by $\mu_{j(i,t),t}$, absorb firm-specific shocks such as a layoff, and allow me to compare workers within the same firm. W_{it} is a vector of worker characteristics such as educational background, gender (from first names or profile pictures), experience since Ph.D., and position types (e.g., engineers vs. scientists).

The coefficient β_k represents the mobility difference between workers who produce $D_{it}(k) = 1$ and those with neither a paper nor a patent, and γ_k represents the gap between workers with Lagged-$D_{it}(k) = 1$ and those without any innovation in the past three years. For workers who are employed by less innovative nontop

firms, the model predictions in Section 2.3 can be translated to:

$$\text{Prediction } 1 \rightarrow \boldsymbol{\beta}_P > 0, \ \boldsymbol{\beta}_{PQ} > 0 \tag{4.2}$$

$$\text{Prediction } 2(a) \rightarrow \boldsymbol{\beta}_{PQ} < \boldsymbol{\beta}_P, \text{ whereas } 2(b) \rightarrow \gamma_{PQ} > \gamma_P \text{ and } \gamma_{PQ} > 0$$

$$\text{Prediction } 3 \rightarrow \gamma_Q > 0$$

Table 3: Effects of Papers & Matched Patents on Job Mobility

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
CS Papers at t : $D_{it}(P)$ vs. $D_{it}(PQ)$						
Paper only	0.0351	-0.0012	0.0052	0.0186	0.0032	0.0018
	(0.0060)	(0.0042)	(0.0024)	(0.0034)	(0.0036)	(0.0009)
Paper+Matched Patent	0.0200	0.0016	-0.0023	0.0135	0.0020	0.0020
	(0.0102)	(0.0062)	(0.0063)	(0.0059)	(0.0055)	(0.0027)
CS Papers in $[t-3, t-1]$: Lagged-$D_{it}(P)$ vs. Lagged-$D_{it}(PQ)$						
Paper only	0.0009	0.0009	0.0077	0.0036	-0.0003	0.0047
	(0.0035)	(0.0031)	(0.0022)	(0.0017)	(0.0028)	(0.0008)
Paper+Matched Patent	0.0195	0.0068	0.0039	0.0107	0.0003	0.0053
	(0.0065)	(0.0051)	(0.0045)	(0.0039)	(0.0047)	(0.0020)
Patents unrelated to CS Papers						
$D_{it}(Q)$	-0.0125	-0.0047	-0.0052	-0.0003	0.0084	0.0022
	(0.0023)	(0.0029)	(0.0039)	(0.0011)	(0.0025)	(0.0013)
Lagged-$D_{it}(Q)$	0.0052	-0.0013	0.0058	0.0023	0.0033	0.0004
	(0.0018)	(0.0024)	(0.0027)	(0.0009)	(0.0021)	(0.0009)
Mean	.1141	.0655	.0746	.0180	.9485	.0067
N	224K	66K	121K	224K	66K	121K
Adj. R^2	.1377	.0179	.1167	.0350	.0112	-.0109

Notes: This table presents regression estimates of equation (4.1). The estimation sample is at Person × Year level, restricted to years with nonmissing full-time employment after PhD. The first three columns show the results for any move between firms as the dependent variable, $M_{it} = 1[j(i, t+1) \neq j(i, t)]$, separately by the group of origin $j(i, t) \in \{\text{Non-top Firms, Top Firms, Academia}\}$. The next three columns have $M_{it} = 1[j(i, t+1) \in \text{Top Firms}]$ as the dependent variable.

All regressions control for education background (whether a bachelor's degree was granted in the United States, and Ph.D. school fixed effect), a cubic polynomial of years since Ph.D. as experience (divided by 10), current position types (scientist/engineer/manager), seniority or academic job rank based on job titles on LinkedIn, and firm-year fixed effects. Standard errors are robust and clustered at (Ph.D. school, graduation cohort) level.

I estimate the model (4.1) separately for workers who are currently employed by nontop firms, top firms, or academia as in Figure 4. For each person, I keep years of full-time employment post Ph.D. and post 2000.[34]

At nontop firms, a worker who has a new paper but no matched patent is 3.5 pp ($t \approx 6$) or 26.3% more likely to move than a coworker with similar experience but no innovation (column 1 of Table 3).[35] Workers who produce a paper with a matched patent, $D_{it}(PQ) = 1$, also see a 2.0 pp or 15.0% increase in next-year inter-firm mobility, but to a lesser extent than workers with $D_{it}(P) = 1$. The positive effects of having any paper for next-year mobility among employees at non-top firms is consistent with Figure 4 and provides evidence for Prediction 1 (4.2) on public learning. The smaller mobility effect of having a paper with a matched patent than of having just a paper is consistent with Prediction 2a, though the gap ($\hat{\beta}_{PQ} - \hat{\beta}_P$) is not significantly different from zero.

There is stronger evidence for asymmetric learning in the mobility responses to lagged innovation outputs, as summarized in 2b (4.2). Holding constant current-period publications, having any paper but not matched patent in the past three years no longer matter for inter-firm mobility even for workers at nontop firms. In contrast, the presence of a paper with a matched patent during $[t-3, t-1]$, denoted by Lagged-$D_{it}(PQ) = 1$, predicts a $\hat{\gamma}_{PQ} = 2.0$-pp or 14% significant increase in job mobility.

Papers in index year t or the past three years do not predict a job movement out

---

[34]There are employment records before 2000 for earlier Ph.D. cohorts but given the relatively short history of CS conferences, I collected publications data post 2000.

[35]Appendix Table C1 shows estimates of Poisson regression:

$$
E[M_{it}|D_{it}, \text{Lagged-}D_{it}, W_{it}, j(i,t)] = exp\left(\sum_{k \in \{P, PQ, Q\}} \beta_k\, D_{it}(k) + \sum_{k \in \{P, PQ, Q\}} \gamma_k\, \text{Lagged-}D_{it}(k) + W'_{it}\Gamma + \mu_{j(i,t),t} \right)
$$

$$(4.3)$$

of top firms, which are often viewed as the top of the job ladder in the tech industry (column 2). Column 3 shows that productive authors in academia experience a 0.5-0.8 pp increase in mobility relative to coworkers without a paper. Whether a paper has a matched patent or not, however, does not matter in academia, where less than 5% of CS papers are filed as patent applications.[36]

Columns 4-6 of Table 3 presents the estimates of the event study model (4.1) defining the outcome of interest as mobility to a top firm in the next year. For workers at nontop firms, I interpret moving to a top firm as moving up the job ladder to a more productive firm, though of course there are some highly productive startups (many of which ultimately are acquired by the top firms). I consider alternative upward mobility measures in Section 4.3. I find a 1.4-1.9 pp increase in upward mobility when employees of nontop firms publish a new paper, consistent with Figure 1b and supporting Prediction 1b. Lagged papers with a matched patent predict another 1 pp or 30% increase in upward mobility, further providing evidence for asymmetric learning from initially private information. For workers in academia (column 6), papers predict moving to a top firm, which may represent a wage increase (Section 4.3). I do not find evidence that CS papers (and matched patents) matter for retention or movement between top firms (column 5). This is consistent with the model predictions that positive signals matter more for workers at nontop firms who receive a larger wage increase on the market than those who are already at the top of the job ladder.[37]

Finally, I show a delayed mobility response to patent applications that are

---

[36]Papers that are filed as patent applications by academics often represent collaborations with the industry and matter less for tenure evaluation within academia.

[37]Employees at top firms may also rather move to a non-research job internally than leave a top firm. I show in Section 4.3 that publications are positively associated with a promotion in top firms, which typically reflect a wage increase, and negatively correlated with becoming an engineer as opposed to scientist within top firms.

unrelated to CS papers, consistent with Prediction 3. Workers who file new patent applications are less likely to leave their incumbent employers in the same year ($\hat{\beta}_Q < 0$ in columns 1-3). But I find a positive relationship between lagged patent applications during $[t-3, t-1]$ and job mobility among workers at nontop firms (columns 1 and 4) in support of (4.2). In comparison with the mobility effects of CS papers, traditional patenting is less predictive of job movements for computer scientists. This feature is not surprising given the emphasis on publications in job ads for computer scientists (Appendix Figure B2).

In summary, computer scientists who publish papers at nontop firms are more likely to move to a new firm and move up the job ladder, providing evidence for public employer learning from CS papers. At nontop firms, workers who produce papers with matched patents, which are initially private information, experience a delayed increase in job mobility. Alternative tests that exploit within-person variation in innovation output provide similar evidence of employer learning (Appendix Table C2). In addition, I show in Appendix Figure C1 that the mobility responses to publications and matched patents are larger for less experienced workers, consistent with the hypothesis that learning is most important at early career stages as has been emphasized in the existing literature (e.g., Lange 2007).

4.3 Additional Evidence of Learning: Wage Growth and Promotions

The evidence of employer learning is not limited to the inter-firm mobility responses above. I summarize additional effects of publications and matched patents on: (1) moving to a higher-wage firm based on the H1-B salary data; (2) moving from industry to academia; (3) within-firm promotions and reallocation according to job titles.

An alternative definition of upward mobility is moving to a higher-paid firm.

31

Without access to administrative wage records, I use the average wages posted for H1-B or PERM workers at firm×year or firm×year×position levels. At nontop firms, workers with a new paper are 2-3 pp more likely to move to a higher-wage firm the next year (Appendix Table C3). Lagged papers with matched patents also increase likelihood of moving to a higher-wage employer, and the likelihood of moving to a higher-wage position, both of which support asymmetric learning as in Prediction 2.

Publishing a paper also increases the chance that workers in the industry move to academia the next year (columns 5-6 of Table C3). Employees who publish at either nontop or top firms are twice as likely to move to an academic employer than their coworkers. Whether the paper has a matched patent application, which may indicate higher commercialization value, matters less for getting a job in academia. The role of publications in helping academia identify talent from the industry is policy relevant, given the rising concerns about the competition for AI talent between academia and the private sector (Gofman and Jin 2022).

Promotions are another set of important mobility outcomes. Pastorino (2023) estimates that employer learning explains 25% of early-career wage growth within a firm, and promotions are responsible for almost all of the impact of learning on wages. I consider a change in job titles as a promotion if the new title includes keywords such as "senior".[38] I estimate (4.1) with internal promotion as the outcome variable on stayers who are not moving to a new firm the next year. CS papers (new or lagged with matched patents) are positively associated with promotions (columns 1-3 of Appendix Table C4). Although incumbent employers can differentiate between $D_{it}(P)$ and $D_{it}(PQ)$ without any delay, I do not find that workers who

---

[38]For example, a change from "engineer" to "senior engineer" or "staff engineer" is coded as a promotion. In academia, getting tenured is a promotion.

produce a paper with a matched patent are getting promoted faster than those with only a paper. This finding supports the promotion-as-signal model in Waldman (1984), which predicts that incumbent employers would delay promotions (public signals) to retain privately known talent longer.

That being said, there is evidence of internal reallocation of workers even under the presence of asymmetric information. Column 4 of Table C4 shows that $D_{it}(PQ)$ workers at nontop firms are 1 pp more likely than $D_{it}(P) = 1$ workers to become a research scientist the next year.[39] In contrast, innovation outputs are less predictive of a worker becoming a manager (columns 8-9), and appear to be negatively correlated with becoming a engineer at top firms (column 7).

5 Structural Analysis

The event-study style analysis above provides evidence of both public and private learning about CS researchers. Assessing the quantitative importance of employer learning for the efficiency of labor allocation, however, requires a structural analysis. I therefore turn to estimate the model in Section 2. To preview the results, I find that learning from on-the-job research matters as much as learning from initial information such as PhD ranking. Disclosing patent applications one year faster has a small but positive impact on overall innovation, which is driven by faster sorting of high-ability workers into top firms.

5.1 Model Estimation

I discuss the structural parameters and present the estimation procedure based on the nested fixed point algorithm (Rust 1987; Rust 1994). The goal is to find

---

[39]Moving from a non-scientist to a scientist role is not coded as a promotion, unless the job title includes keyword such as "senior".

estimates that maximize the joint likelihood of job histories and innovation outputs of computer scientists in the first five years post PhD.[40][41] I show the model fit and the learning process evaluated at the maximum likelihood estimates.

5.1.1 Parameters and Identification

Table 4 provides an overview of the main ingredients of the model, which are parameters that govern: (1) prior belief conditional on initial information, (2) labor supply, (3) firm productivity, and (4) worker productivity.

Table 4: Overview of Model Parameters

Parameters	Description
I. Common Prior	
δ	Given initial information I_{i1}, prior:

$$Pr(\alpha_i = H | I_{i1}; \delta) = \frac{exp\left(\delta' X(I_{i1})\right)}{1 + exp\left(\delta' X(I_{i1})\right)} \qquad (5.1)$$

Parameters	Description
II. Labor Supply	
$b, \{\rho_G\}, \{\eta_{\cdot G}\}$	Worker's utility (2.6): weight on log wage and GEV-preferences (2.4)
$\{\lambda\}, \{\Lambda\}$	Prob. re-entering the labor market (2.5) and moving bet. academia and industry.
III. Firm Productivity	
$\overline{\phi}_j$	Baseline productivity of 16 groups of employers (Appendix Table D2)
$\{\phi_j(P), \phi_j(Q)\}$	Returns to each type of innovation: patent $\phi_j(Q)$ calibrated, $\phi_j(PQ)$ is assumed to be a weighted avg of $\phi_j(P)$ and $\phi_j(Q)$
IV. Worker Productivity	
$p_\alpha, \widetilde{p}_\alpha, q_\alpha$	Ability-specific productivity in innovation (Table 1)

First, I assume employers form a common prior based on initial information

---

[40]The first few years are particularly important for employer learning (Altonji and Pierret 2001; Farber and Gibbons 1996; Lange 2007; Pallais 2014). I also show in Appendix Figure C1 that the mobility responses to innovation signals are higher for less experienced workers.

[41]I assume employer belief is held fixed after $t = 5$. A simple terminal value is defined as the discounted sum of rents from a worker, given that she is expected to stay $= 1/\lambda(I_{i5})$ years beyond $t = 5$ (see equation 7.4 in Appendix A1).

I_{i1} about a new Ph.D., comprising (i) the rank of their PhD institution, (ii) the number of papers published during graduate school, and (iii) the type of their first job (i.e., the "nest" G of their first employer). This information is combined with coefficients δ in equation (5.1) to specify the prior probability that a given individual has high productivity.[42]

Second, the parameters of the labor supply model defined in (2.7) include the utility weight b on log wage, parameters in the GEV-distributed preferences ρ_G, and additional parameters η_G that enter the worker's utility for all employers in nest G.[43] The ratio $\frac{b}{\rho_G}$, the probability of being on the job market λ's, and the exogenous rate Λ's at which academia is open to industry employees and vice versa, jointly enter the labor supply elasticity (2.9) that determines the markdown of wages relative to the expected revenue product of labor. These parameters are identified by revealed preferences and variations in retention rates within and between nests.

There are more than seven thousand unique employers in the balanced panel of workers. Following Bonhomme et al. (2022), I classify them into 16 meta-firms and estimate meta-firm-specific productivity.[44]. The classification is shown in Appendix Table D2, and follows the nested structure of the labor supply model (equation 2.6). Specifically, I divide jobs in academia into four meta firms based on academic ranking: two tenure track (nest $G = 1$) and 2 post-doc / non-tenure track employers ($G = 2$). The third nest comprises the top 6 tech companies, each as a meta-firm on its own. Finally, I group the non-top employers into 6 meta-firms,

---

[42]Any information observed by employers but not by me is assumed to be absorbed by the initial nest of a person's first job, G_{i1}. Let r_i denote the rank of PhD school, n_i denote the number of papers before PhD. I define a vector of controls: $X(I_{i1}) = (r_i, r_i^2, n_i, n_i^2, 1[G_{i1} = \text{Tenure-track}], 1[G_{i1} = \text{Postdoc}], 1[G_{i1} = \text{Top firms}], 1[G_{i1} = \text{Nontop firms}])$.

[43]The nested logit choice probability is fully specified in equation (7.1) in Appendix A1.

[44]This grouping is equivalent to assuming that employers within each meta-firm are perfect substitutes to workers, i.e. diversity between employers within a meta-firm is not valued, as remarked in Dixit and Stiglitz (1977).

grouped by their patenting activity. Each meta firm (henceforth j) has 4 productivity parameters as shown in equation (2.1): a baseline productivity $\bar{\phi}_j$, a return to research publication $\phi_j(P)$ that is estimated, a return to patents unrelated to publications $\phi_j(Q)$ that is calibrated, and a return to publications with matched patents $\phi_j(PQ)$ as a weighted average of $\phi_j(P)$ and $\phi_j(Q)$. [45] The baseline productivity $\bar{\phi}_j$ matters for the average wage and thus the size of a meta-firm. $\phi_j(P)$ matters for the allocation to publication-oriented tasks, and is identified from movers who become more (less) likely to publish when moving to a higher (lower) $\phi_j(P)$ employer.

The fourth set of parameters represents the ability-specific productivity in Table 1. Conditional on information, coworkers who have different abilities would be assigned the same innovation task (τ). The gap in their publication rate allows me to identify the ability-specific publication rates, p_H versus p_L.

5.1.2 Estimation Procedure

Denote by Γ the free parameters of the model in Table 4 other than the δ parameters of the prior (5.1). Given data on the each person's employment by meta firms $\{j(i,t)\}$ and outputs $d_{it} := [d_{it}(P), d_{it}(PQ), d_{it}(Q)]$, I search for estimates of

---

[45]I estimate a regression of any patent application on firm fixed effects and worker characteristics. I rank nontop firms according to the estimated fixed effects, which are also used to calibrate $\phi_j(Q)$ (Table D2). I calibrate $\phi_j(PQ) = 2.25 \times \phi_j(P) + 0.25 \times \phi_j(Q)$, which fits the data better than an unweighted average.

(δ, Γ) that solve:

$$max_{(\delta, \Gamma)} \, ln \left(\prod_i L_i(\{j(i,t), d_{it}\} \,|I_{i1}; \boldsymbol{\delta}, \boldsymbol{\Gamma}) \right) \qquad (5.2)$$

$$= \sum_i ln \left(\sum_\alpha \underbrace{Pr(\alpha|I_{i1}; \boldsymbol{\delta})}_{\text{prior}} \times L_i(\{j(i,t), d_{it}\} \,|I_{i1}, \boldsymbol{\alpha}; \boldsymbol{\Gamma}) \right)$$

in which $L_i(\cdot|I_{i1}, \boldsymbol{\alpha}; \, \boldsymbol{\Gamma}) = \prod_t \underbrace{s_{itj(i,t)}(I_{it}, \widetilde{I}_{it}; \boldsymbol{\Gamma})}_{\text{labor supply}} \times \underbrace{Pr(D_{it} = d_{it}|\boldsymbol{\alpha}, \boldsymbol{\tau}_{itj(i,t)}; \boldsymbol{\Gamma})}_{\text{innovation output}}$

in which information evolves according to (2.3), and the unobserved ability α_i is treated as a random effect. The equilibrium labor supply is solved as the fixed point given a guess for Γ. $\{\tau_{itj(i,t)}\}$ are the task allocations in equilibrium. Following Rust (1987), I use a nested fixed point algorithm with three steps:

Step 0. Given a guess of δ on initial information, form the prior (5.1) shared by employers.

Step 1. Given a guess of Γ, solve each employer's problem backward from $t = T$: at every possible information state $(I, \widetilde{I})$, calculate the labor supply $\{s_{tj}\}$ given the wages posted by firms, and iterate until I reach the fixed point $s_{tj}(I, \widetilde{I}; \Gamma)$: $s_{tj}(I, \widetilde{I}; \Gamma) = s \left(w \left(s_{tj}(I, \widetilde{I}; \Gamma) \right) \right)$.[46]

Step 2. Find the maximum likelihood estimates that solve (5.2). Parameters δ on initial information and other free parameters Γ are jointly estimated as in econometric frameworks with unobserved heterogeneity (e.g., Card and Hyslop 2005, Wooldridge 2005).

5.1.3 Estimation Results

I estimate the model on a balanced panel of 18,860 workers who graduated between 2005 and 2018 and have employment records in the first five years post

---

[46]See Proposition 1 for the existence of the fixed point in equation (2.16).

PhD.[47] The predicted share of workers at each employer, found as the fixed point (2.16) given the maximum-likelihood estimates for Γ (Table D3), falls roughly on the 45-degree line that matches with the actual shares, as shown in Figure 5.

Figure 5: Model Fit: Allocation of Workers across Employers, $\hat{s}_{tj}$ vs. s_{tj}

(a) $t = 1$ (b) $t = 5$
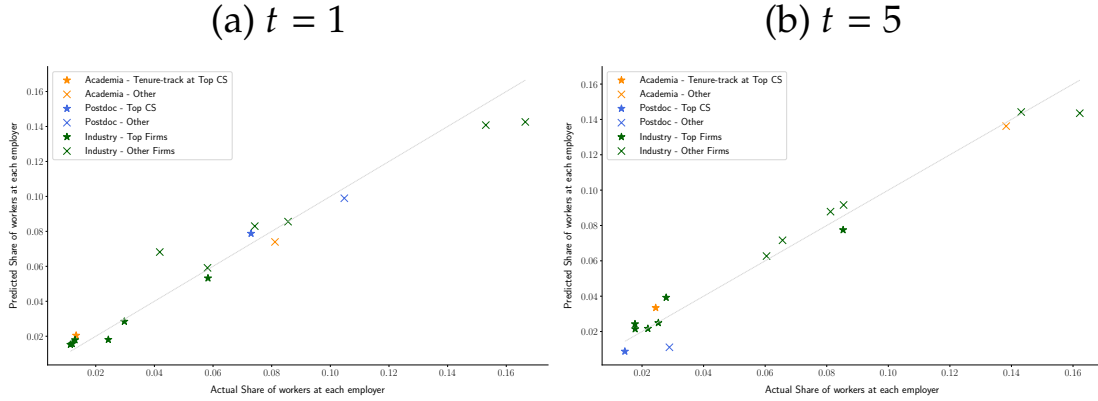
Notes: This figure shows the predicted share of workers at each meta-firm $\hat{s}_{tj}$ against the actual share s_{tj}, at $t = 1$ and $t = 5$. Given the estimated parameters in Table D1, I forward simulate the employment path and innovation outputs by each worker in the balanced sample, holding fixed initial information including the initial nest at $t = 1$ (see 5.1). In the simulated sample, I compute $\hat{s}_{tj}$ as the share of workers employed by j, at experience t (yrs after PhD).

High-ability workers are estimated to be four times as likely to produce a paper per unit of time on innovation tasks as the L-ability. Conditional on publishing a paper, H are more than twice as likely to have a patent application matched to the paper, which indicates a higher quality innovation. H is also more likely to produce patents unrelated to papers than L, but the relative gap in patenting is smaller than in research publications.[48]

Figure 6 displays the distribution of employer beliefs, separately for the top 10% cited computer scientists (as a proxy for H) versus the bottom 90%. At $t = 1$ beliefs about these two groups overlap substantially. But employers appear to tell

---

[47]This sample is comparable to the full sample that I use to test for employer learning in Section 5 (see Table D3).

[48]These estimates validate the assumptions $p_H > p_L, \widetilde{p}_H > \widetilde{p}_L, q_H > q_L$ under which model predictions are derived.

them apart quickly based on their research outputs after PhD. At $t = 5$, there is an obvious divergence of beliefs about the bottom 90% versus the top 10%. Comparing the gap between posterior and prior across PhD institutions, I find that the market overestimates the ability of graduates from the top 10 programs (see Appendix Figure D1).

Figure 6: Histograms of Employer Beliefs: Top 10% Computer Scientists versus Others

(a) $t = 1$ (b) $t = 5$
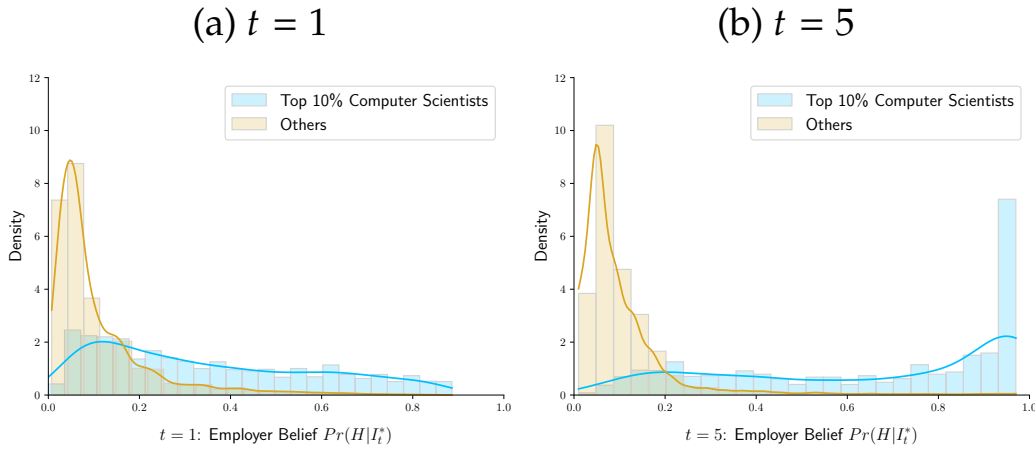
Notes: I rank computer scientists by their cumulative citations and total number of papers and patent applications five years after PhD.

Tenure-track employers have the highest returns to research publications ($\phi_j(P)$ in Table D2) and assign more innovation tasks given any employer belief (Figure D2). Top firms (except for Apple) on average have higher returns to research papers and assign more innovation tasks than nontop firms. The gap between top and nontop firms in innovation tasks is larger for workers with employer belief in the range of $[0.20, 0.50]$, who have a nontrivial chance of being H-ability but are not fully discovered yet. Such workers would be better off at a more productive firm that provides more research opportunities.

5.2 Impacts of Employer Learning on Allocative Efficiency

Given the estimated model, I assess the impact of employer learning on the efficiency of labor allocation. To do so, I consider five mechanisms that matter for workers' sorting between employers and task allocations within a firm:

1. Employer learning from patents unrelated to papers, $D_{it}(Q)$;
2. Employer learning from publications with a matched patent, $D_{it}(PQ)$;
3. Employer learning from research publications, $D_{it}(P) + D_{it}(PQ)$;
4. Initial sorting between nests, G_{i1};
5. Access to initial information $I_{i1} \setminus G_{i1}$.

The first three mechanisms capture employer learning from on-the-job research outputs after Ph.D., while the last two concern any initial information observed by employers that shape the common prior about each worker and her sorting between academia and industry at $t = 1$. Each mechanism can change the allocation of labor by affecting the evolution of employer beliefs.[49]

I measure the efficiency of talent allocation by the mean publication rate of computer scientists, an outcome that is shaped by task allocation within each firm and sorting of workers between firms. Figure 7(a) shows how this outcome changes when I shut down the mechanisms one by one (in the order above), relative to the benchmark where all mechanisms are at play.[50] Shutting down learning from patent applications unrelated to papers reduces publication rate by just 0.9%, which makes sense as H and L are not as different in patenting as in producing papers.

---

[49]For example, if employers do not update their beliefs based on research publications, they would not assign additional publication-oriented tasks internally to employees who publish, and publication authors would also not receive higher wage offers from other firms than their coworkers without a paper.

[50]The benchmark model I estimated takes all five mechanisms into account. Given $(\hat{\delta}^{MLE}, \hat{\Gamma}^{MLE})$ in Table D1, I first forward-simulate the employment path and innovation output of workers without shutting down any mechanism. The benchmark publication rate on the simulated sample is 9.23%, similar to the mean observed in the estimation sample. In each counterfactual, I re-simulate the data and compute the mean publication rate under the alternative set of mechanisms.

The first substantial drop in publication rate occurs when I shut down employer learning from CS publications. That is, employers no longer update their beliefs based on whether workers have published. As a result, employers do not assign more publication-oriented tasks to successful authors, nor are H-ability sorted into more productive firms as efficiently as before. Together, employer learning from innovation outputs $\{(D_{it}(P), D_{it}(PQ), D_{it}(Q))\}$ accounts for the 15.8% of the overall publication rate.[51] Appendix Figure D3 further shows that top firms and academic employers experience suffer more when they do not learn from the innovation output of workers.

Figure 7: Decomposition of Publication Rate (Efficiency of Talent Allocation)
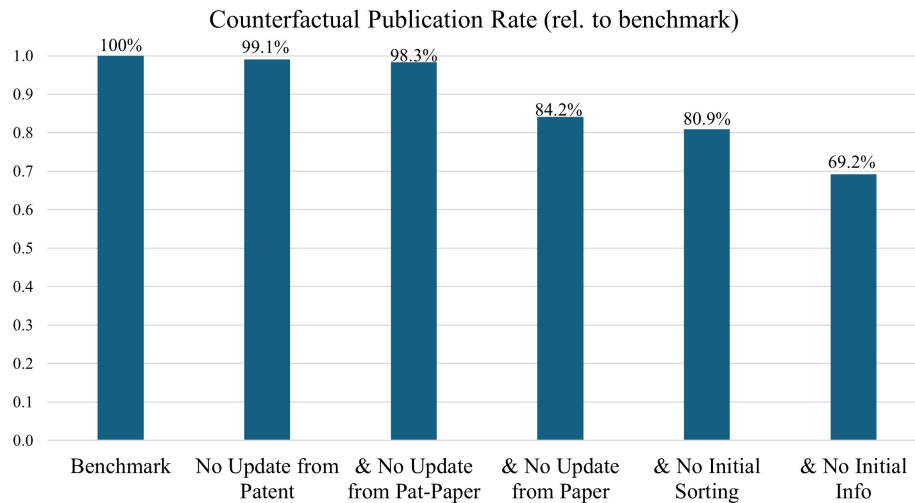
Shutting down initial sorting between nests further reduces the publication rate by 3.3%, while other initial information such as PhD school and papers before PhD accounts for 11.7%. When all five learning mechanisms are removed, the 69.2% of publications remained are explained purely by firm heterogeneity and

---

[51]Appendix D4 provides a between-within decomposition and shows that 30% of the effect is driven by between-firm sorting whereas the rest is explained by less efficient task alloctaion within firms.

worker heterogeneity.[52]

5.2.1 Shapley Value of Each Learning Mechanism

To address concerns that the counterfactual results in Figure 7 are shaped by the order of the mechanisms, I estimate the average marginal impact of each mechanism on allocative efficiency a la Shapley (1953).[53] I compute the counterfactual publication outcome under 2^5 possible sets of mechanisms.[54] The Shapley value of mechanism $m \in \mathbb{M} = \{1, 2, 3, 4, 5\}$ is:

$$SV_m = \sum_{S \subseteq \mathbb{M} \setminus \{m\}} \frac{|S|! \times (|M| - |S| - 1)!}{|M|!} \times \underbrace{\left(E\left[p_{it} | S \cup \{m\}; \hat{\delta}, \hat{\Gamma}\right] - E\left[p_{it} | S; ; \hat{\delta}, \hat{\Gamma}\right] \right)}_{\text{Change in publication rate when adding mechanism } m}$$

(5.3)

As shown in Table 5, the five mechanisms jointly account for 31% of the publication rate, consistent with the last bar in Figure 7. I normalize the Shapley values of the five features such that they sum to one. The most important feature is employer learning from the presence of CS papers, with a normalized Shapley value of 49.9%.[55] Initial information ranks second with an explanatory power of 40.2%. Initial sorting based on information seen by employers but not by me

---

[52]H remains more productive in innovation than L, but without employer learning from either initial information or subsequent outputs, H and L are assigned the same amount of innovation task within each firm and move between firms at the same rates.

[53]Shapley (1953) has been applied to attribute model prediction or goodness-of-fit to individual features (e.g., Grömping 2007, Lindeman and Gold 1980). Huneeus, Kroft, and Lim (2021) also uses Shapley value to decompose the variance of earning inequality on multiple sources of variation in their counterfactual analysis.

[54]For example, if the set of mechanisms included is $\{1, 2, 3\}$, the counterfactual data generation allows employers to update beliefs based on innovation outputs after PhD, but the initial prior about every worker equals to the mean prior 0.13 fitted on the original data. If the set is empty, only worker ability and firm heterogeneity matter for innovation outcome.

[55]Since the model does not consider self-selection by workers, this estimate may be interpreted as an upper bound of the role of employer learning in efficient allocation of labor.

matters less with a mere value of 2%. Learning from patent applications which are not disclosed immediately to outside employers explains the remaining 8%, which are smaller than learning from publications but nonnegligible.

Table 5: Shapley Values of Employer Learning vs. Initial Conditions

	Employer Learning			Initial Conditions	
	Patent	Paper-Patent	Paper	Initial Sorting	Initial Info
Impact on Mean Publication Rate					
SV_m (5.3)	0.0014	0.0009	0.0142	0.0005	0.0114
Pct Change	1.51%	0.98%	15.39%	0.55%	12.40%
Normalized SV_m	4.89%	3.18%	49.92%	1.78%	40.23%

Notes: This table shows the estimated Shapley value of each mechanism, the percentage change relative to the benchmark outcome when all five mechanisms are considered, and the normalized value such that they sum to one.

5.3 Asymmetric Learning on Efficiency

Would reducing asymmetric information improve the efficiency of talent allocation? On one hand, increasing public information about workers can expedite positive assortative matching between firms and workers. On the other hand, allocation to ability-revealing tasks, like training, would be inefficiently lower when current firms lose their information rents (e.g., Acemoglu and Pischke 1998).

I answer this question by considering a "symmetric" counterfactual where $D_{it}(P)$, $D_{it}(PQ)$, and $D_{it}(Q)$ are disclosed simultaneously.[56] Given the estimates in Table D1, I forward simulate the employment path and research production by workers, holding fixed the initial information. For the counterfactual, I begin with the same set of workers, and find the equilibrium contracts set by employers under simultaneous information disclosure.

---

[56]This scenario is possible if the patents office discloses patent applications immediately, or if workers can announce patent applications themselves without worrying about a non-disclosure agreement.

Figure 8: Upward Mobility from Non-top to Top Firms, Asymmetric vs. Symmetric (dashed)
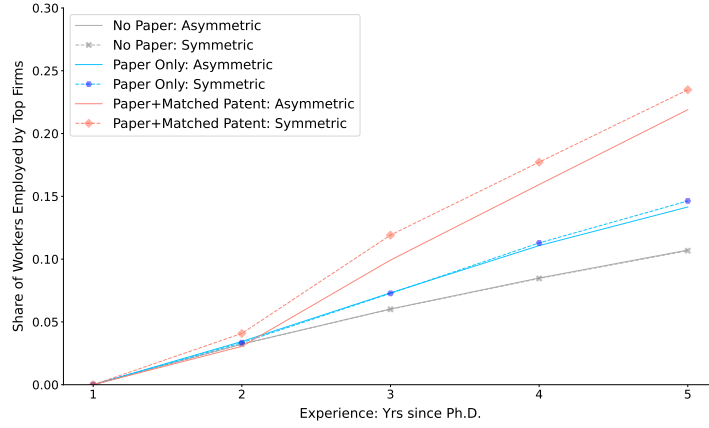
Figure 8 displays the upward mobility for workers who start at nontop firms but produce different innovation output, under the asymmetric benchmark versus the symmetric counterfactual. Workers with output $D_{it}(PQ) = 1$ can be told apart immediately from workers with a paper only $D_{it}(P) = 1$ under the counterfactual. Relative to the asymmetric benchmark, they move to top firms more quickly. In contrast, workers who only have a paper, $D_{it}(P) = 1$, or no paper at all are as likely to move upward as before.[57]

The overall publication rate is 1% higher under symmetric disclosure (Table 6). Top firms experience a 5.5% increase in innovation output, while nontop firms see a 2.4% decrease. The change in innovation outcomes comes from two sources: 1) faster sorting of productive workers from nontop to top firms, and 2) changes in within-firm allocation to publication-oriented tasks. To decompose the change, I hold fixed the task allocations made by firms under asymmetric learning in the simulations for the symmetric counterfactual. The last set of results in Table 6

---

[57] $D_{it}(P) = 1$ workers from non-top firms are still more likely to be H-ability than those with no publication. They may produce papers with a matched patent later on and benefit from the reduction of asymmetric information, which would explain the small increase in the share employed by the top at $t = 5$ in this group.

Table 6: Innovation Output under Asymmetric vs. Symmetric Learning

	Paper		Paper-Patent	
	Mean	% Change	Mean	% Change
Asymmetric Benchmark				
Overall	0.0923		0.0176	
Top Firms	0.1120		0.0390	
Nontop Firms	0.0429		0.0131	
Symmetric				
Overall	0.0931	0.97%	0.0179	1.39%
Top Firms	0.1181	5.50%	0.0412	5.50%
Nontop Firms	0.0419	-2.37%	0.0128	-2.12%
Symmetric, τ| Asymmetric				
Overall	0.0935	1.30%	0.0180	2.07%
Top Firms	0.1179	5.33%	0.0416	6.69%
Nontop Firms	0.0422	-1.55%	0.0130	-0.73%

shows that CS publications would increase by 1.3% rather than 0.97% if employers were not to assign tasks differently in response to the reduction of asymmetric information. Faster positive assortative matching therefore accounts for 134% of the increase in publication rate when there is simultaneous disclosure of papers and patents. Incumbent employers assign fewer innovation tasks just like they would reduce training when they have less monopsony power, countering the efficiency gains from sorting.

6 Conclusion

This paper provides empirical evidence of employer learning and quantifies its role in improving the allocation of labor in the labor market for computer scientists. By building a database that combines the employment histories of newly minted Ph.D.'s in computer science with their publications and patents, I offer more direct tests of public and private employer learning than what has been

shown in the employer learning literature.

Publishing a CS conference proceeding increases the inter-firm mobility of a worker at nontop firm by 30%, and almost doubles her chance of moving to one of the top-6 tech firms in the following year. This pattern suggests a strong role for public employer learning in the reallocation of workers between firms. Higher-quality papers often coincide with a closely related patent application, but the application remains private for 18 months. Using an event-study style analysis, I show that authors of such papers experience a delayed increase in job mobility, which provides evidence for asymmetric learning.

The mobility changes around the publications versus (matched) patent applications are consistent with predictions from the dynamic framework of employer learning and sorting in this paper, in which I introduce information frictions about talent into a model of imperfect labor market competition. I estimate a structural version of the model and find that in the absence of employer learning from public research records, the innovation output of early-career computer scientists would drop by 16%. Disclosing patent applications one year faster would increase innovation by 1%, driven by a faster rate of positive assortative matching.

A limitation about the data is that CS Ph.D.'s on LinkedIn are more likely to work in industry than in academia. I show that workers who publish papers in the industry are also more likely to move to academia, which suggests those publications are also valued by academic employers. But more data needs to be collected on academic computer scientists to investigate if encouraging tech firms to participate in CS conferences reduces the AI brain drain from academia (e.g., Jurowetzki et al. 2021).

The dynamic framework of employer learning can also be adapted to quantify the role of employer learning on reallocation of labor in other labor markets. This

paper suggests that even for a high-skilled group with strong credentials like PhD computer scientists, information frictions are prevalent and result in substantial misallocation of workers between and within firms. It is likely that reducing information frictions about workers would matter even more for the matching between firms and workers in lower-skilled or less transparent labor markets.

References

ACEMOGLU, D. AND J.-S. PISCHKE (1998): "Why Do Firms Train? Theory and Evidence," *The Quarterly Journal of Economics*, 113, 79–119.

AIPA (1999): "American Inventors Protection Act, 35 U.S.C." .

ALTONJI, J. G. AND C. R. PIERRET (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

BECKER, G. (1964): *Human Capital*, University of Chicago Press.

BONHOMME, S., T. LAMADON, AND E. MANRESA (2022): "Discretizing Unobserved Heterogeneity," *Econometrica*, 90, 625–643.

BOOZER, M. (1994): *Wages and Turnover with Imperfect Information and Imperfect Competition*, Doctoral Dissertation, Princeton University.

BURDETT, K. AND D. T. MORTENSEN (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273.

CARD, D., A. R. CARDOSO, J. HEINING, AND P. KLINE (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36, S13–S70.

CARD, D. AND D. R. HYSLOP (2005): "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica*, 73, 1723–1770.

DIXIT, A. K. AND J. E. STIGLITZ (1977): "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 67, 297–308.

FARBER, H. S. AND R. GIBBONS (1996): "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111, 1007–1047.

GIBBONS, R. AND L. KATZ (1991): "Layoffs and Lemons," *Journal of Labor Economics*, 9, 351–380.

GOFMAN, M. AND Z. JIN (2022): "Artificial intelligence, human capital, and innovation," *Journal of Finance, Forthcoming*.

GREENWALD, B. C. (1986): "Adverse Selection in the Labour Market," *Review of Economic Studies*, 53, 325–347.

GRÖMPING, U. (2007): "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition," *The American Statistician*, 61, 139–147.

HAGER, S., C. SCHWARZ, AND F. WALDINGER (2023): "Measuring Science: Performance Metrics and the Allocation of Talent," *Working Paper*.

HENDRICKS, K. AND R. H. PORTER (1988): "An Empirical Study of an Auction with Asymmetric Information," *The American Economic Review*, 865–883.

HUNEEUS, F., K. KROFT, AND K. LIM (2021): "Earnings Inequality in Production Networks," Tech. rep., National Bureau of Economic Research.

IMBENS, G. AND J. WOOLDRIDGE (2007): *What's New in Econometrics Lecture 11: Discrete Choice Models*, NBER.

JUROWETZKI, R., D. HAIN, J. MATEOS-GARCIA, AND K. STATHOULOPOULOS (2021): "The Privatization of AI Research(-ers): Causes and Potential Consequences - From university-industry interaction to public research brain-drain?" *arXiv 2102.01648*.

KAHN, L. B. (2013): "Asymmetric Information between Employers," *American Economic Journal: Applied Economics*, 5, 165–205.

KAHN, L. B. AND F. LANGE (2014): "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *The Review of Economic Studies*, 81, 1575–1613.

LANGE, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 1–35.

LI, J. (2013): "Job Mobility, Wage Dispersion, and Technological Change: An Asymmetric Information Perspective," *European Economic Review*, 60, 105–126.

LINDEMAN, R. H., M. P. F. AND R. Z. GOLD (1980): *Introduction to Bivariate and Multivariate Analysis*, Glenview, IL: Scott, Foresman.

LIU, D. C. AND J. NOCEDAL (1989): "On the limited memory BFGS method for large scale optimization," *Mathematical programming*, 45, 503–528.

MANNING, A. (2003): *Human Capital and Training*, Princeton University Press, 301–322.

MCFADDEN, D. (1973): "Conditional logit analysis of qualitative choice behavior," in *Fontiers in Econometrics*, ed. by P. Zarembka, New York: Academic press, 105–142.

MILGROM, P. AND S. OSTER (1987): "Job Discrimination, Market Forces, and the Invisibility Hypothesis," *The Quarterly Journal of Economics*, 102, 453–476.

PALLAIS, A. (2014): "Inefficient Hiring in Entry-Level Labor Markets," *American Economic Review*, 104, 3565–99.

PASTORINO, E. (2023): "Careers in Firms: The Role of Learning about Ability and Human Capital Acquisition," *Journal of Political Economy*, Forthcoming.

POSTEL-VINAY, F. AND J.-M. ROBIN (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

——— (1994): *Estimation of dynamic structural models, problems and prospects: discrete decision processes*, Cambridge University Press, 119–170, Econometric Society Monographs.

SARSONS, H. (2017): "Recognition for Group Work: Gender Differences in Academia," *American Economic Review*, 107.

SCHÖNBERG, U. (2007): "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25, 651–691.

SHAPLEY, L. S. (1953): *A Value for n-Person Games*, Princeton: Princeton University Press, 307–318.

STERN, S. (2004): "Do Scientists Pay to Be Scientists?" *Management Science*, 50, 835–853.

STEVENS, M. (1994): "A Theoretical Model of On-the-Job Training with Imperfect Competition," *Oxford Economic Papers*, 46, 537–62.

TERVIÖ, M. (2009): "Superstars and Mediocrities: Market Failure in the Discovery of Talent," *The Review of Economic Studies*, 76, 829–850.

WALDMAN, M. (1984): "Job Assignments, Signalling, and Efficiency," *The RAND Journal of Economics*, 15, 255–267.

WOOLDRIDGE, J. M. (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 20, 39–54.

7 Appendix

A. Proofs and Model Extension

A0. Model Timeline

There are $T \geq 3$ discrete periods in this model. At least three periods are needed to fully capture the information revelation process: innovation is produced at an initial employer at $t = 1$; the presence of a paper ($D_{it}(P) + D_{it}(PQ) \in \{0, 1\}$) is known by the beginning of $t = 2$; whether a paper from $t = 1$ has a matched patent application ($D_{it}(PQ)$ vs. $D_{it}(P)$), and whether there is a patent unrelated to paper ($D_{it}(Q)$), are not revealed until $t = 3$.

1. ($t = 1$) New PhD graduates enter the labor market.

   (a) Given initial information $\{I_{i1}\}$ about workers, employers post wages $\{w_{i1j}\}$ simultaneously and choose the share of time each worker can spend on innovation tasks, $\tau_{i1j} \in [0, 1]$.

   (b) Each worker observes the wages posted by all firms and chooses an initial employer $j(i, 1)$ that maximizes her utility (2.6) at $t = 1$.

   (c) Innovation outputs (Figure 2) are realized by the end of $t = 1$ and are fully known to i's incumbent employer $j(i, 1)$. There are 6 possible values of the vector of indicators:

   $$(D_{it}(P), D_{it}(PQ), D_{it}(Q)) \in \{(1, 0, 1), (1, 0, 0), (0, 1, 1), (0, 1, 0), (0, 0, 1), (0, 0, 0)\}$$

2. ($t = 2$) Public information I_{i2} and private information $\widetilde{I}_{i2}$ at the beginning of $t = 2$ evolve according to (2.3)

   (a) Firms update their beliefs about a worker's research ability, post new $\{w_{i2j}\}$ simultaneously, and choose task allocation $\{\tau_{i2j}\}$. A firm's problem is summarized in (2.10) and (2.13).

   (b) Workers re-enter the labor market with probability (2.5). If they are on the market, they observe contracts posted by potential employers, draws new idiosyncratic preferences that are independent from her preferences at $t = 1$, and solve (2.6) again. Otherwise, they stay at their original employers, $j(i, 2) = j(i, 1)$.

   (c) Repeat 1(c).

3. ($t = 3$) Public information I_{i3} and private information $\widetilde{I}_{i3}$ at the beginning of $t = 3$ evolve according to (2.3):

   $$\text{Public } I_{i3} = I_{i2} \cup \widetilde{I}_{i2} \cup \{D_{i2}(P) + D_{i2}(PQ)\}$$
   $$\text{Private } \widetilde{I}_{i3} = \{(D_{i2}(P), D_{i2}(PQ), D_{i2}(Q))\}$$

   Repeat the rest of 2.

4. ($t > 3$) Repeat 3 until period T after which the model concludes.

A1. Backward Induction

Details on Workers' Problem (Section 2.2.2)
Workers who are on the market can choose a new employer as discussed in Section 3.2.1 (see equation 2.6). The choice of an employer is summarized by a static nested logit model. Given a choice set C, workers on the market draw idiosyncratic preferences $\{\epsilon_{itj}\}$ from a GEV distribution (2.4).[58]

Given the wages posted by firms $\{w_j\}$, define the inclusive value of a nest G of employers as:

$$W_G := ln\left(\sum_{j \in G} exp(\frac{b}{\rho_G} ln(w_j))\right)$$

Therefore, the choice probabilities given public belief $\pi = Pr(\alpha_i = H|I_{it})$ that enter the labor supply can be written as:

$$s_{j|C} = \underbrace{s_{j|G(j)}}_{\text{choose } j \in G(j)} \times \underbrace{s_{G(j)|C}}_{\text{choose nest } G(j) \in C} \tag{7.1}$$

$$\forall G : s_{G|C} = 1[G \in C] \times \frac{exp(\eta_G(\pi) + \rho_G \times W_G)}{\sum_{G' \in C} exp(\eta_{G'}(\pi) + \rho_{G'} \times W_{G'})}$$

$$\forall j \in G : s_{j|G} = \frac{exp(\frac{b}{\rho_G} ln(w_j))}{exp(W_G)}$$

Backward Induction:

I solve for the subgame perfect MPBNE in Definition 1 via backward induction.

---

[58] Workers who have entered the industry may not be as likely to receive academic offers as workers who have been working in academia. We assume the choice set C includes all nests at $t = 1$ for new PhDs. At $t > 1$, C includes academic nests (tenure-track or postdocs) for industry employees with probability Λ_{JA}, and $C = \{$Nontop Firms, Top Firms$\}$ with probability $(1 - \Lambda_{JA})$. Similarly, for workers in academia at $t > 1$, industry employers are in the choice set C with probability Λ_{AJ}. We take $(\Lambda_{AJ}, \Lambda_{JA})$ as model parameters that are estimated in Section 5.

Last Period $t = T$

At the last period T, employer j's value function is the sum of expected revenue generated by period-T employees net wages:

$$V_{Tj} = \underbrace{\sum_{i: j(i,T-1)=j} v_{Tj}^{(1)}(I_{iT}, \widetilde{I}_{iT})}_{\text{Incumbent}} + \underbrace{\sum_{i: j(i,T-1)\neq j} v_{Tj}^{(0)}(I_{iT})}_{\text{Workers Outside}} \tag{7.2}$$

where I_{iT} represents the public information about worker i at the beginning of T, while $\widetilde{I}_{iT}$ represents the private information known only if worker i is an incumbent employee. Employers derive optimal contracts for incumbent versus new workers separately, due to differences in their labor supply and information about their ability.

Incumbent Employees

Given information $(I_{iT}, \widetilde{I}_{iT})$ about an incumbent employee i, employer j solves:

$$v_{Tj}^{(1)}(I_{iT}, \widetilde{I}_{iT}) = max_{w,\tau} \; \underbrace{s_j^{(1)}(w, w_{-j}; I_{iT})}_{\text{labor supply}} \times \underbrace{\left(E_{\alpha|I_{it} \cup \widetilde{I}_{it}}[Y_j(\alpha, \tau)] - w \right)}_{\text{MRPL net wage}} \tag{7.3}$$

$$\text{where } s_j^{(1)}(w, w_{-j}; I_{iT}) = \underbrace{1 - \lambda(I_{iT})}_{\text{off market}} + \underbrace{\lambda(I_{iT}) \times E_C[s_{j|C}(w, w_{-j})]}_{\text{on market \& enter j again}}$$

where w_{-j} are wages posted by other employers given public information I_{iT}, taken as given by the j.[59] Public information I_{iT} matters for the probability at which the worker re-enters the labor market (2.5). This formulation is equivalent to defining a terminal value function $v_{(T+1)j}$ as follows, assuming that employers no longer update their beliefs based on innovation output from $t = T$ onward. Employers will continue to pay (w, τ) set in T until a worker leaves the firm. Given the rate at which workers re-enter the labor market per period (λ's in 2.5), the expected tenure of worker i in firm j after T is approximately $\frac{1}{\lambda(I_{iT})}$. Define

$$v_{(T+1)j}(w, \tau) := \sum_{k=0}^{1/\lambda(I_{iT})} \beta^k \left(E[Y_j(\alpha, \tau)] - w \right) \tag{7.4}$$

$$v_{Tj}^{(1)} := max_{w,\tau} \; s_{Tj}^{(1)} \times \left(E[Y_j(\alpha, \tau)] - w + \beta \times v_{(T+1)j}(w, \tau) \right)$$

$$= s_{Tj}^{(1)} \times \left(E[Y_j(\alpha, \tau)] - w \right) \times \frac{1 - \beta^{1/\lambda(I_{iT})}}{1 - \beta}$$

The terminal value function defined above will yield the same optimal contract as the one

---

[59]Wages are posted simultaneously by employers. In equilibrium, $w_{-j} = w_{-j}^{(0)}(I_{iT})$, the optimal wages outside firms would post given public information I_{iT}.

that maximizes (7.3), since $\frac{1-\beta^{1/\lambda(I_{iT})}}{1-\beta}$ is independent of the choice variables (w, τ).

Take derivatives of the objective function (7.3) over wage w:

$$\frac{\partial s_j^{(1)}(\boldsymbol{w}, w_{-j}; I_{iT})}{\partial w} \times \left(E_{\alpha | I_{iT} \cup \widetilde{I}_{iT}}[Y_j(\alpha, \boldsymbol{\tau})] - w \right) - s_j^{(1)}(\boldsymbol{w}, w_{-j}; I_{iT}) = 0 \qquad (7.5)$$

letting $G = G(j)$,

$$\frac{\partial s_j^{(1)}(\boldsymbol{w}, w_{-j}; I_{iT})}{\partial w} = \lambda(I_{iT}) \times \left(\underbrace{\frac{\partial s_{j|G}}{\partial w} \times E_C[s_{G|C}]}_{(a)} + \underbrace{s_{j|G} \times \frac{\partial E_C[s_{G|C}]}{\partial w}}_{(b)} \right)$$

$$(a) = \frac{b/\rho_G}{w} \times s_{j|G} \times (1 - s_{j|G})$$

$$(b) = \frac{b}{w} \times s_{j|G} \times E_C[s_{G|C} \times (1 - s_{G|C})]$$

Merging the equations above yields the labor supply elasticity w.r.t. wage for the incumbent worker i:

$$\xi_{iTj}^{(1)} := \frac{\partial ln(s_j^{(1)}(\boldsymbol{w}, w_{-j}; I_{iT}))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C[\underbrace{\frac{\lambda_G \times s_{j|G} \times s_{G|C}}{s_j^{(1)}}}_{(c)} \times \underbrace{(1 - \rho_G \, s_{j|G} \, s_{G|C} - (1 - \rho_G)s_{j|G})]}_{(d)}$$

$$(7.6)$$

where (c) represents the ratio of the probability of an incumbent worker getting on the market and choosing j again to the probability of staying at j. This ratio converges to 1 when $\lambda \to 1$ (that is, incumbent employees search for new jobs with probability 1). On the other hand, when λ is small, the labor supply of incumbent workers is highly inelastic. Wages at T would be 0 if $\lambda = 0$. If the choice set includes all employers and $\rho_G = 1$, (d) can be reduced to $(1 - s_j)$.

Plugging $\xi_{iTj}^{(1)}$ into the first order condition (7.5), the optimal wage for an incumbent worker is:

$$w_{iTj}^{(1)} = \boldsymbol{w}_{Tj}^{(1)}(w_{-j}; I_{iT}, \widetilde{I}_{iT}) = E_{\alpha | I_{iT} \cup \widetilde{I}_{iT}}[Y_j(\alpha, \tau_{iTj}^{(1)})] \times \underbrace{\xi_{iTj}^{(1)} \times \left(1 + \xi_{iTj}^{(1)} \right)^{-1}}_{\text{markdown}} \qquad (7.7)$$

In equilibrium (Definition 1), $\forall I: w_{-j} = w_{-j}^{(0)}(I)$, and we have $w_{Tj}^{(1)}(I, \widetilde{I}) = \boldsymbol{w}_{Tj}^{(1)}(w_{-j}(I); I, \widetilde{I})$.

Taking the derivative of (7.3) over allocation to publication-oriented tasks, τ,

$$\underbrace{\frac{\partial s_j^{(1)}}{\partial \tau}}_{=0} + s_j^{(1)} \times \frac{\partial E_{\alpha | I_{iT} \cup \widetilde{I}_{iT}}[Y_j(\alpha, \tau)]}{\partial \tau} \geq 0 \tag{7.8}$$

$$\text{define } \tau_{Tj}^*(I, \widetilde{I}) := \frac{1}{\zeta} E_{\alpha | I \cup \widetilde{I}} \left[p_\alpha \widetilde{p}_\alpha \phi_j(PQ) + p_\alpha (1 - \widetilde{p}_\alpha) \phi_j(P) - 1 \right]$$

$$\longrightarrow \tau_{iTj}^{(1)} = max\{0, min\{1, \tau_{Tj}^*(I_{iT}, \widetilde{I}_{iT})\}\}$$

<u>*Outside Workers*</u>

For an outside worker i from $j(i, T-1) \neq j$, employer j only has access to public information I_{iT}. The value function is therefore expected over private information $\widetilde{I}$ conditional on I_{iT}. Specifically, employer j solves:

$$v_{Tj}^{(0)}(I_{iT}) = max_{w,\tau} E_{\widetilde{I}|I_{iT}} \left[\underbrace{s_j^{(0)}(w, w_{-j}; I_{iT}, \widetilde{I})}_{\text{labor supply}} \times \underbrace{\left(E_{\alpha|I_{iT} \cup \widetilde{I}}[Y_j(\alpha, \tau)] - w \right)}_{\text{MRPL net wage}} \right] \tag{7.9}$$

$$\text{where } s_j^{(0)}(w, w_{-j}; I_{iT}, \widetilde{I}) = \lambda(I_{iT}) \times E_C \left[s_{j|C}(w, w_{(-j)}; I_{iT}, \widetilde{I}) \right]$$

w_{-j} are wages posted by other employers. Since $(-j)$ includes the incumbent employer $j(i, T-1)$ that has private information about this worker, w_{-j} and therefore $s_j^{(0)}$ in (7.9) varies by private information $\widetilde{I}$.

Taking the derivative of (7.9) over the wage w posted by j:

$$E_{\widetilde{I}|I_{iT}} \left[\frac{\partial s_j^{(0)}}{\partial w} \times \left(E_{\alpha|I_{iT} \cup \widetilde{I}}[Y_j(\alpha, \tau_j^{(0)})] - w \right) - s_j^{(0)}(w, w_{-j}; I_{iT}, \widetilde{I}) \right] = 0 \tag{7.10}$$

Conditional on the not-yet-known $\widetilde{I}$:

$$\frac{\partial s_j^{(0)}(w, w_{-j}; I_{iT}, \widetilde{I})}{\partial w} = \lambda(I_{iT}) \times \left(\underbrace{\frac{\partial s_{j|G}}{\partial w} \times E_C[s_{G|C}]}_{(e)} + s_{j|G} \times \underbrace{\frac{\partial E_C[s_{G|C}]}{\partial w}}_{(f)} \right)$$

$$(e) = \frac{b/\rho_G}{w} \times s_{j|G} \times (1 - s_{j|G})$$

$$(f) = \frac{b}{w} \times s_{j|G} \times E_C[s_{G|C} \times (1 - s_{G|C})]$$

Merging the equations above yields the labor supply elasticity w.r.t. wage for new workers:

$$\xi_{iTj}^{(0)}(\widetilde{I}) := \frac{\partial ln(s_j^{(0)}(\boldsymbol{w}, w_{-j}; I_{iT}, \widetilde{I}))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C\left[\frac{s_{G|C}}{E_C[s_{G|C}]} \times \left(1 - \rho_G\, s_{j|G}\, s_{G|C} - (1-\rho_G)s_{j|G}\right)\right] \tag{7.11}$$

where $\widetilde{I}$ matters for the wages set by the incumbent employer of this outside worker and thus each choice probability in (7.11). In contrast with the elasticity $\xi_{iTj}^{(1)}$ of an incumbent worker (7.6), $\lambda(I_{iT})$, the probability getting on the market, no longer matters for the elasticity to a new employer j.[60] In addition, j is uncertain about $\widetilde{I}$ and the elasticity is specific to $\widetilde{I}$ conditional on public information I_{iT}. Plugging the above into FOC (7.10), the optimal wage for outside employee i can be written as:[61]

$$w_{iTj}^{(0)} = \left(1 + E_{\widetilde{I}|I_{iT}}\left[\frac{s_j^{(0)}}{E_{\widetilde{I}|I_{iT}}[s_j^{(0)}]} \times \xi_{iTj}^{(0)}(\widetilde{I})\right]\right)^{-1} \tag{7.12}$$

$$\times E_{\widetilde{I}|I_{iT}}\left[\frac{s_j^{(0)}}{E_{\widetilde{I}|I_{iT}}[s_j^{(0)}]} \times \xi_{iTj}^{(0)}(\widetilde{I}) \times E_{\alpha|I_{iT}\cup\widetilde{I}}[Y_j(\alpha, \tau)]\right]$$

Taking the derivative of (7.9) over task allocation τ,

$$\frac{\partial E_{\widetilde{I}|I_{iT}}[s_j^{(0)} \times E_{\alpha|I_{iT}\cup\widetilde{I}}[Y_j(\alpha, \tau)]]}{\partial \tau} \geq 0 \tag{7.13}$$

$$\rightarrow E_{\widetilde{I}|I_{iT}}[s_j^{(0)} \times (-1 + E_{\alpha|I\cup\widetilde{I}}[\phi_j(PQ)p_\alpha\widetilde{p}_\alpha + \phi_j(P)p_\alpha(1-\widetilde{p}_\alpha)] - \zeta\tau)] \geq 0$$

$$\rightarrow \tau_{iTj}^{(0)} = E_{\widetilde{I}|I_{iT}}\left[\frac{s_j^{(0)}}{E_{\widetilde{I}|I_{iT}}[s_j^{(0)}]} \times \tau_{Tj}^*(I_{iT}, \widetilde{I})\right]$$

that is, the task allocation for an outside worker is a weighted average of what firm j would have set if $\widetilde{I}$ is known. The weight on $\widetilde{I}$ equals the likelihood of $\widetilde{I}$ conditional on public I_{iT} and the case that the worker moves to j.

---

[60]New workers are predicted to be paid a higher wage than equally productive incumbent workers, due to their more elastic labor supply when $\lambda < 1$. In this paper I do not have data on wages and thus do not test this prediction.

[61]At information state $(I, \widetilde{I})$, $s_j^{(0)} = s_j^{(0)}(w_{iTj}^{(0)}, w_{-j}; I, \widetilde{I})$, which equals to $s_j^{(0)}(I, \widetilde{I})$ in equilibrium, evaluated at $w_{iTj}^{(0)} = w_{Tj}^{(0)}(I)$ and $w_{-j}(I, \widetilde{I})$ (see Definition 1)

Middle Periods $t = 2, ..., (T-1)$

$$V_{tj}\left(\bigcup_{\text{worker } i} I_{itj}\right) = \underbrace{\sum_{i:\, j(i,t-1)=j} v_{tj}^{(1)}(I_{it} \cup \widetilde{I}_{it})}_{\text{Incumbent}} + \underbrace{\sum_{i:\, j(i,t-1)\neq j} v_{tj}^{(0)}(I_{it})}_{\text{Workers Outside}} \qquad (7.14)$$

Employer j solves the following for incumbent workers:

$$v_{tj}^{(1)}(I_{it}, \widetilde{I}_{it}) = max_{w,\tau}\ \underbrace{s_j^{(1)}(w, w_{-j};\, I_{it})}_{\text{expected labor supply}} \times \underbrace{\left(E_{\alpha|I_{it}\cup\widetilde{I}_{it}}[Y_j(\alpha, \tau)] + \beta\, E_D[v_{(t+1)j}^{(1)}(I', \widetilde{I}')\,|\,\tau] - w\right)}_{\text{MRPL at t \& discounted continuation value, net wage}}$$

$$(7.15)$$

Employers now take into the expected continuation value from stayers at $(t+1)$:

$$E_D[v_{(t+1)j}^{(1)}(I', \widetilde{I}')\,|\,\tau] = \sum_D Pr(D|I_{it} \cup \widetilde{I}_{it}, \tau) \times v_{(t+1)j}^{(1)}(I'(D), \widetilde{I}'(D)) \qquad (7.16)$$

$$\text{in which } D = (D_{it}(P), D_{it}(PQ), D_{it}(Q))$$

$$Pr(D|I_{it} \cup \widetilde{I}_{it}, \tau) = \sum_\alpha Pr(\alpha|I_{it} \cup \widetilde{I}_{it}) \times Pr(D|\alpha, \tau) \text{ as in Table 1}$$

$$I'(D) = I_{it} \cup \widetilde{I}_{it} \cup \{D_{it}(PQ) + D_{it}(P)\}$$

$$\widetilde{I}'(D) = \{D\}$$

The optimal wages at $t < T$, as shown in (2.11) and repeated below, can be derived the same way as wages at $t = T$:

$$w_{itj}^{(1)} = \left(E_{\alpha|I_{it}\cup\widetilde{I}_{it}}[Y_j(\alpha, \tau)] + \beta\, E[v_{(t+1)j}^{(1)}(I', \widetilde{I}')\,|\,\tau_{itj}^{(1)}]\right) \times \underbrace{\xi_{itj}^{(1)} \times \left(1 + \xi_{itj}^{(1)}\right)^{-1}}_{\text{markdown}} \qquad (7.17)$$

The firm-specific labor supply elasticity of an incumbent worker or a new worker can be written the same as equations (7.6) (7.11), respectively. The difference in wages at $t < T$ from wages at $t = T$ is that employers also share some of the expected continuation value with the worker (marked down by the inverse of labor supply elasticity). In other words, the dynamic monopsonistic wages in this framework are front-loaded. Once a worker has entered the firm, wages for incumbent employees are lower unless they keep re-entering the labor market ($\lambda \to 1$). The gap between an incumbent and equally productive new worker may be interpreted as a signing bonus or stock options contracted upon entry.

Optimal task allocations now depend on the changes to continuation value given

innovation outputs:

$$\tau_{itj}^{(1)} = max\{0, min\{1, \ \tau_{tj}^*(I_{it}, \widetilde{I}_{it})\}\} \tag{7.18}$$

$$\tau_{tj}^*(I, \widetilde{I}) := \underbrace{\frac{1}{\zeta} \times E_{\alpha|I \cup \widetilde{I}}\left[-1 + \sum_{k \in \{P, PQ, Q\}} \phi_j(k) \times \frac{\partial E[D_{it}(k)|\alpha, \tau]}{\partial \tau}\right]}_{\text{return to innovation today}} + \frac{\beta/\bar{\phi}_j}{\zeta} \times \underbrace{\frac{\partial E[v_{(t+1)j}^{(1)}(I', \widetilde{I}')|\tau]}{\partial \tau}}_{\text{change in value from stayer}}$$

$$\tag{7.19}$$

in which the dynamic return to assigning more publication-oriented tasks today:

$$\frac{\partial E[v_{(t+1)j}^{(1)}(I', \widetilde{I}')|\tau]}{\partial \tau} \tag{7.20}$$

$$= \sum_{D} \frac{\partial Pr(D|I_{it} \cup \widetilde{I}_{it}, \tau)}{\partial \tau} \times v_{(t+1)j}^{(1)}(I'(D), \widetilde{I}'(D))$$

$$= \sum_{\alpha} Pr(\alpha|I_{it} \cup \widetilde{I}_{it}) \times \sum_{k \in \{P, PQ\}} \underbrace{\frac{\partial Pr(D_{it}(k) = 1|\alpha, \tau)}{\partial \tau}}_{\text{see Table 1}} \times \underbrace{\left(E[v_{(t+1)j}^{(1)}|\alpha, \text{Paper}] - E[v_{(t+1)j}^{(1)}|\alpha, \text{No Paper}]\right)}_{(*)}$$

where $(*)$ is the change in the firm's continuation value when employee i produces a paper $(D_{it}(PQ) + D_{it}(P) = 1)$ versus not, expected over other patenting activity, $D_{it}(Q)$, which does not vary by τ (Table 1).

The optimal contracts for a new worker maximize (2.13). The derivation is similar to that of $t = T$, and the solutions are presented in Section 2.2.3.

In summary, we have derived the optimal wages as expressed in (2.11,2.14), and the optimal task allocations in (2.12,2.15). In equilibrium, employers set wages and allocate workers to publication-oriented tasks, conditional on information about workers and taking as given the contracts set by other employers. The expected labor supply from incumbent employees and from new workers are determined by the wages set by potential employers.

First Period $t = 1$

New PhD's are on the market at $t = 1$ and observe the contracts posted by all employers. Firms simultaneously solve the following conditional on common initial information

I_{i1}:

$$V_{1j}\left(\bigcup_i I_{i1}\right) = \sum_i v_{1j}(I_{i1}) \tag{7.21}$$

$$v_{1j}(I_{i1}) = max_{\boldsymbol{w},\boldsymbol{\tau}} \underbrace{s_j(\boldsymbol{w}, \boldsymbol{w}_{-j}; I_{i1})}_{\text{labor supply}} \times \left(\underbrace{E_{\alpha|I_{i1}}[Y_j(\alpha, \boldsymbol{\tau})]}_{\text{MRPL at } t=1} + \underbrace{\beta \times E[v_{2j}^{(1)}(I', \widetilde{I}')|\boldsymbol{\tau}]}_{\text{continuation value}} - \boldsymbol{w} \right)$$

The FOC for initial wage:

$$\frac{\partial s_j(\boldsymbol{w}, \boldsymbol{w}_{-j}; I_{i1})}{\partial w} \times \left(E_{\alpha|I_{i1}}[Y_j(\alpha, \boldsymbol{\tau})] - w\right) - s_j(\boldsymbol{w}, \boldsymbol{w}_{-j}; I_{i1}) = 0$$
$$\text{where } s_j(\boldsymbol{w}, \boldsymbol{w}_{-j}; I_{i1}) = s_{j|G} \times s_G$$

The elasticity of labor supply to firm $j \in G$ at $t = 1$ equals:

$$\xi_{i1j} = \frac{b}{\rho_G} \times \left(1 - (1 - \rho_G)s_{j|G} - \rho_G s_j\right) \tag{7.22}$$

The optimal contract can then be written as:

$$w_{i1j} = \left(E_{\alpha|I_{i1}}[Y_j(\alpha, \boldsymbol{\tau})] + \beta\, E[v_{2j}^{(1)}(I', \widetilde{I}')|\tau_{i1j}]\right) \times \xi_{i1j} \times \left(1 + \xi_{i1j}\right)^{-1} \tag{7.23}$$

$$\tau_{i1j} = max\{0, min\{1, \frac{1}{\zeta}E_{\alpha|I_{i1}}[p_\alpha \widetilde{p}_\alpha \phi_j(PQ) + p_\alpha(1 - \widetilde{p}_\alpha)\phi_j(P) - 1 + \frac{\beta}{\bar{\phi}_j} \times \frac{\partial E[v_{2j}^{(1)}(I)|\tau]}{\partial \tau}]\}\}$$

where wage markdown equals the inverse of labor supply elasticity in (7.22), and the continuation value changes in τ as in (7.20).

The backward induction from $t = T$ to $t = 1$ is complete.

A2. Model Equilibrium

Proposition 1 - MPBNE under Imperfect Labor Market Competition

Proposition 1 (Existence and Uniqueness of MPBNE) *There exists a strategy profile $\{(w_{tj}, \tau_{tj})\}$ that satisfies Definition 1. The equilibrium wages are unique up to a positive scaling factor, and they result in a unique allocation of workers between firms at each possible information state:*

$$s_{tj}(I, \widetilde{I}) = \begin{cases} s_{1j}(w_1(I)) & t = 1 \\ s_{tj}^{(1)}\left(w_{tj}(I, \widetilde{I}), w_{t(-j)}(I)\right) & t > 1, j = j(i, t-1) \\ s_{tj}^{(1)}\left(w_{tj}(I), w_{t(-j)}(I, ,\widetilde{I})\right) & t > 1, j \neq j(i, t-1) \end{cases}$$

which satisfies (2.16).

Proof:

Existence: I solve the game between firms by backward induction in Appendix A1, starting from period T. In the last period, the wages are marked down from the expected flow profit by the inverse of labor supply elasticity. I show in 2.9 that the firm-specific elasticity is a continuous function of the probabilities of workers choosing each firm, which in turn are continuous functions of wages. Since $s_T \circ w_T : [0,1] \to [0,1]$ is a continuous function, by Brouwer fixed-point theorem, there exists a fixed point s_T^* such that $s_T^* = s_T(w_T(s_T^*))$.

At $t = 1, 2..., T-1$, firms post contracts with knowledge of the equilibrium that will be played from $(t+1)$ onwards. Since workers are impatient and solve a repeated static problem (2.6), only wages matter for the allocation of workers between firms, denoted by s_t^*, for all possible information states at t. By the same argument at T, s_t^* is a fixed point to the function $s_t \circ w_t$.

Stacking them together, $s^* = (s_1^*, s_2^*, ..., s_T^*)$ is a fixed point of $s \circ w$, which establishes the existence of the MPBNE.

Uniqueness: Since workers do not value task allocation τ's when choosing an employer, the task allocation strategies denoted by τ_{tj} are solved to maximize the firm's own expected profit plus discounted continuation value, independently from the decisions by other firms.

Every public information state I is on the equilibrium path due to the fact that workers hold idiosyncratic preference over employers. That is a positive probability that a worker of any public belief π is employed by a firm j in each period. The simplifying assumptions on the workers' problem help avoid multiple equilibria in this game. The equilibrium wages are unique up to scaling by a positive constant, and the equilibrium allocation of workers between firms s^* is unique. $\qquad\square$

Proposition 2 - Equilibrium under Perfect Labor Market Competition

Suppose that the labor supply is perfectly elastic in each period ($\frac{b}{\rho} \to \infty$ and $\lambda \equiv 1$), and that the information is incomplete but symmetric among employers. Under such assumptions, the decision to allocate workers to publication-oriented tasks is equivalent to the decision to provide general skill training that is transferable between firms. We get the familiar result in Becker (1964) that workers who are not credit-constrained bear all costs of training and are paid their full marginal product of labor.

Proposition 2 (Equilibrium under Public Information & Perfect Competition) *If the labor market is perfectly competitive and information is always symmetric, each firm j offers a worker with public information I at the beginning of any period t:*

$$w_{tj}(I) = E_{\alpha|I}[Y_j(\alpha, \tau_{tj}(I))] \tag{7.24}$$
$$in \ which \ \tau_{tj}(I) = argmax_{\tau \in [0,1]} E_{\alpha|I}[Y_j(\alpha, \tau)]$$

Proof of Proposition 2:

The labor supply w.r.t. wage is perfectly elastic under the assumptions that $\frac{b}{\rho_G} \to \infty$ and $\lambda \equiv 1$. Plugging $\xi^{(1)} \to \infty$ (2.9 into the incumbent wage at $t = T$, I have $w_{Tj}^{(1)}(I) =$

$E_{\alpha|I}[Y_j(\alpha, \tau_{Tj}^{(1)}(I))]$. Incumbent workers are paid the full expected value conditional on public information I. There is no dynamic rent for employers at $(T-1)$. The wage in intermediary periods, as shown in (2.11), also equals to the expected value from a worker without leaving any dynamic rent to an employer.

Information is assumed to be symmetric between employers. The wages for outside workers also equal to $E_{\alpha|I}[Y_j(\alpha, \tau_{Tj}^{(1)}(I))]$. Since the continuation value equals zero at all employers, allocating workers to publication-oriented tasks also becomes a static decision:

$$\tau_{tj}(I) = max\{0, min\{1, \frac{1}{\zeta}E_I[-1 + \sum_k \phi_j(k) \times \frac{\partial E[D_{it}(k)|\alpha, \tau]}{\partial \tau}]\}\}$$

The costs of publication-oriented tasks are fully deducted from workers' wages (see 2.1). That is, workers are bearing all costs of innovation. They are not credit constrained as they earn a positive wage from routine tasks. The choices of publication-oriented tasks would be first best in each period, just like the choice of general skill training made by workers who are not credit constrained in Becker (1964). □

If the labor market is perfectly competitive but information is asymmetric as in (2.3), less informed employers face a problem similar to Hendricks and Porter (1988) and would adopt a mixed strategy to randomize their wage bids (Boozer 1994; Li 2013). Otherwise, there is always adverse selection (Greenwald (1986)). It is unclear, however, if incumbent employers would allocate workers to publication-oriented tasks efficiently.

A3. Model Predictions

Derivation of Prediction 1: Mobility in Response to Publications

Conditional on (prior) public information I, let π_1 denote the posterior belief that a worker is high-ability when she publishes a paper, and π_0 denote the posterior belief when she does not:

$$\pi_1 = Pr(\alpha = H|I, D(P) + D(PQ) = 1) = \frac{p_H \times Pr(H|I)}{p_H \times Pr(H|I) + p_L \times Pr(L|I)}$$

$$\pi_0 = Pr(\alpha = H|I, D(P) + D(PQ) = 0) = \frac{(1 - p_H \times \tau) \times Pr(H|I)}{(1 - p_H \times \tau) \times Pr(H|I) + (1 - p_L \times \tau) \times Pr(L|I)}$$

where task allocation τ in equilibrium is set optimally by her employer given the information it has. Under the assumption that high-ability workers are more likely to publish, i.e. $p_H > p_L$, we have $\forall \tau \in (0, 1]: \pi_1 > \pi_0$.

a) The probability that the workers stays at her employer j varies by the posterior public belief about her ability:

$$s_j^{(1)}(\pi_1) = 1 - \lambda(\pi_1) \times (1 - E_C[s_{j|C}(\pi_1)])$$

$$s_j^{(1)}(\pi_0) = 1 - \lambda(\pi_0) \times (1 - E_C[s_{j|C}(\pi_0)])$$

The difference between which represents the gap in turnover when a worker publishes a paper:

$$s_j^{(1)}(\pi_1) - s_j^{(1)}(\pi_0) = \underbrace{(\lambda(\pi_0) - \lambda(\pi_1)) \times}_{\leq 0 \text{ at nontop firms}} (1 - E_C[s_{j|C}(\pi_0)])$$

$$+ \lambda(\pi_1) \times \underbrace{(E_C[s_{j|C}(\pi_1)] - E_C[s_{j|C}(\pi_0)])}_{\text{diff in choosing j again if on market}}$$

Under A2, workers with more positive public belief from non-top firms are at least as likely to be on the market as workers with lower employer belief: $\pi_1 > \pi_0 \rightarrow \lambda(\pi_1) \geq \lambda(\pi_0)$. The sign of the second term is determined by if a π_1-worker is more likely to choose the same firm again if she is on the job market.

Under the nested-logit structure, the choice probability for the worker with public belief π is:

$$E_C[s_{j|C}(\pi)] = s_{j|G(j)}(\pi) \times E_C[s_{G(j)|C}(\pi)]$$
$$= \frac{exp(\frac{b}{\rho_G}ln(w_j(\pi)))}{exp(W_G(\pi))} \times E_C[\frac{exp(\eta_{G(j)}(\pi) + W_{G(j)}(\pi))}{\sum_{G \in C} \eta_G(\pi) + W_G(\pi)}]$$

Equilibrium wages at all firms are non-decreasing in π, which suggests $\forall G : \frac{\partial}{\partial \pi}W_G(\pi) \geq 0$ for option value within nest G (**??**). We have $E_C[s_{j|C}(\pi_1)] < E_C[s_{j|C}(\pi_0)]$ unless the wage increase at firm j is disproportionately higher than the increase in option value at $G(j)$ and at other nests of employers. This exception would not happen at a lower-productivity firm that is less sensitive to worker research ability than more productive counterparts.

$$\lambda(\pi_0) \leq \lambda(\pi_1) \,\&\, E_C[s_{j|C}(\pi_0)] < E_C[s_{j|C}(\pi_1)] \rightarrow s_j^{(1)}(\pi_1) < s_j^{(1)}(\pi_0)$$

That is, conditional on prior, workers who publishes a paper are more likely to leave a nontop/less productive firm than coworkers without a publication.

b) Conditional on re-entering the job market, π_1-workers are more likely to choose firms that have higher returns to publications. Firms with higher $\phi_j(P)$ or $\phi_j(PQ)$ assign more publication-oriented tasks:

$$\frac{\partial^2 \tau_j}{\partial \phi_j(P\cdot)\partial\pi} \geq 0$$

which implies the production of research publications is supermodular in equilibrium. Since wages are increasing in the expected returns to innovation, we also have

$$\frac{\partial^2 w_j}{\partial \phi_j(P\cdot)\,\partial\pi} \geq 0$$

The disproportionately higher wage increase at a more productive firm (higher $\phi_j(P)$ or $\phi_j(PQ)$) will attract workers with higher market belief:

$$ln\left(\frac{s_{j|G(j)}(\pi_1)}{s_{j|G(j)}(\pi_0)}\right) = \frac{b}{\rho_{G(j)}} \times ln\left(\frac{w_j(\pi_1)}{w_j(\pi_0)}\right) - \left(W_{G(j)}(\pi_1) - W_{G(j)}(\pi_0)\right)$$

$$> 0 \text{ at firms with high returns to publications}$$

The positive assortative matching matters for marginal workers who would not have spent as much time on innovation task without the positive signal. If π_1, π_0 are significantly high, the worker may spend 100% of time on research at any firm, in which case there is no sorting as in a standard AKM framework.[62]

Derivation of Prediction 2: Job Mobility under Asymmetric Information $D_{it}(PQ)$ vs. $D_{it}(P)$

Consider two workers A, B from firm j with the same public information I at the beginning of period t. The incumbent employer observes that the publication by worker A is accompanied by a patent, but the publication by B is not: $D_{A(t-1)}(PQ) = 1 > D_{B(t-1)}(PQ)$, while outside employers only observe that both workers produce a publication at $(t-1)$. Let $\pi = Pr(\alpha = H|I)$ denote the public belief about A and B. The private belief held by employer j about worker A, $\widetilde{\pi}_A = Pr(\alpha = H|I, D_{A(t-1)}(PQ) = 1) = \frac{\pi \times p_H^*}{\pi \times p_H^* + (1-\pi)p_L^*}$, is higher than the private belief about B under the assumption that $p_H^* > p_L^*$.

a) Based on the labor supply in equation (2.8), the difference between A and B in the probability of staying at their employer j at t is:

$$s_{tj}^{(1)}(I, \widetilde{I}_A) - s_{tj}^{(1)}(I, \widetilde{I}_B) = \underbrace{\lambda(I)}_{>0} \times \underbrace{\left(E_C[s_{j|C}(\widetilde{\pi}_A)] - E_C[s_{j|C}(\widetilde{\pi}_B)]\right)}_{(*)} \qquad (7.25)$$

$$\widetilde{\pi}_A > \widetilde{\pi}_B \rightarrow (*) > 0$$

Given the same public information, the two workers are equally likely to be on the job market. Equilibrium wages from outside employers would be equal for A and B. However, employer j with private information would set a higher wage for the A, resulting in higher retention of A: $s_{tj}^{(1)}(I, \widetilde{I}_A) > s_{tj}^{(1)}(I, \widetilde{I}_B)$.

---

[62]If the wages are set in a AKM fashion as follows, there is no sorting between high π and more productive firms

$$\forall \pi : ln(w_j(\pi)) = \alpha(\pi) + \phi_j$$

$$ln\left(\frac{s_j(\pi_1)}{s_j(\pi_0)}\right) = \frac{b}{\rho_{G(j)}}(\alpha(\pi_1) - \alpha(\pi_0)) - \frac{b}{\rho_{G(j)}}(\alpha(\pi_1) - \alpha(\pi_0)) = 0$$

b) Under the information structure, private information $\widetilde{I}_A, \widetilde{I}_B$ based on their outputs during $(t-1)$ are revealed by the beginning of $(t+1)$. As the market receives more positive signals about worker A than B, we have the new market belief $\pi_A > \pi_B$. Prediction 1 re-applies. Workers with a newly revealed patent matched to their paper at less productive firms experience an increase in inter-firm and upward mobility relative to coworkers with only a paper.

Derivation of Prediction 3: Job Mobility under Asymmetric Information $D_{it}(Q)$

Prediction 3 can be thought of as a corollary of Prediction 2. Patents that are unrelated to papers, denoted by Q, is also revealed with a one-period delay. Since high-ability workers are also more likely to patent $q_H > q_L$, the indicator for any Q is a positive signal about worker ability that is held private by the employer. As a result, workers with $D_{it}(Q)$ are also more likely to stay at the original employer when Q is private, and become more likely to move away from a less productive firm and move up the job ladder Q is public information.

A4. Model Extension - Forward-looking Workers

The benchmark model shuts down the dynamic incentives of workers by assuming that workers choose employers based on wages and idiosyncratic preferences in a given period (2.6). Here I present an extension with forward-looking workers, who consider the option value of re-entering the job market when they choose an employer. Working for a firm that assigns more publication-oriented tasks would be more appealing to a high-ability individual, who can improve the future market belief about her by publishing a paper today.

As before, workers who are on the market hold GEV-distributed preferences over firms (2.4). Those who are not on the market stay with their original employers and hold the preference fixed too. The option value of a worker at information state $(I, \widetilde{I})$ at the beginning of conditional on starting at firm j with preference ϵ_{tj} is:

$$\Omega_{(t+1)j}(I, \widetilde{I}; \epsilon_{tj}) = (1 - \lambda(I)) \times \underbrace{\left(b \, ln(w^{(1)}_{(t+1)j}(I, \widetilde{I})) + \rho_{G(j)}\epsilon_{tj} \right)}_{\text{value if staying at j, holding fixed } \epsilon_{tj}}$$

$$+ \lambda(I) \times E_C \underbrace{\left[ln \left(\sum_{G \in C} exp(\eta_G(I) + W_G(I, \widetilde{I})) \right) \right]}_{\text{option value if re-entering the job market}}$$

Let β_W denote the discount factor of workers. Workers who are on the market observe posted contracts (w_j, τ_j) and choose an employer as follows:

$$j(i, t) = argmax_j \, b \times ln(w_j) + \beta_W \times \underbrace{E_D[\Omega_{(t+1)j}(I, \widetilde{I})|\tau_j, I_{it} \cup \widetilde{I}_{it}]}_{\text{expected option value}} + \rho_G \epsilon_{itj} \qquad (7.26)$$

63

The option value can be transformed into a preference for publication-oriented tasks τ:

$$E_D[\Omega(I, \widetilde{I})|\tau_j, I_{it} \cup \widetilde{I}_{it}] = \sum_D Pr(\boldsymbol{D}|\tau_j, I_{it} \cup \widetilde{I}_{it}) \times \Omega_{(t+1)j}(I(\boldsymbol{D}), \widetilde{I}(\boldsymbol{D})) \qquad (7.27)$$

in which the likelihood of producing $\boldsymbol{D} = (D_{it}(P), D_{it}(PQ), D_{it}(Q))$ by the law of iterated expectations is: $Pr(\boldsymbol{D}|\tau_j, I_{it} \cup \widetilde{I}_{it}) = \sum_\alpha Pr(\alpha|I_{it} \cup \widetilde{I}_{it}) \times Pr(\boldsymbol{D}|\tau_j, \alpha)$.[63]

Given the same likelihood of innovation $Pr(\boldsymbol{D}|\tau_j, \alpha)$ as shown in Table 1, the expected option value can be interpreted as the worker's utility from publication-oriented tasks τ_j. The utility from τ is increasing in the worker's belief about her ability based on current information $(I_{it}, \widetilde{I}_{it})$. That is, workers who are more likely to be high-ability would value publication-oriented tasks more than others.[64]

The expected labor supply (2.8) also depends on the task allocation:

$$\frac{\partial s_{tj}}{\partial \tau_j} \neq 0, \frac{\partial^2 s_{tj}}{\partial \tau_j \partial \pi^*} \geq 0 \text{ for } \pi^* := Pr(H|I_t \cup \widetilde{I}_t)$$

This complicates the optimal contracts set by forward-looking firms. Firms could assign higher τ while keeping wages lower to attract equally capable workers. Workers with a more positive belief that they are high-ability would also be willing to accept a lower wage in exchange of higher τ that can increase their option value on the market, the same as postdocs who "pay to do science" in Stern (2004).

The contracts set by employers in equilibrium are no longer unique. Despite the presence of self-selection by forward-looking workers, firms continue to face a dynamic trade-off when assigning publication-oriented tasks as emphasized in the benchmark model (e.g., equation 2.12).

Prediction 1 on the increased job mobility and upward mobility from less productive firms when workers publish continues to hold. Under asymmetric information, workers with knowledge of $(I, \widetilde{I})$ may stay at incumbent firms to enjoy a higher τ without requiring a higher wage, or voluntarily accept a wage cut at firms that assign higher τ. But the increase in job mobility when $\widetilde{I}$ becomes public information, in Prediction 2(b) and Prediction 3, continue to hold.

---

[63]For simplicity, I assume that conditional on task τ, the transition probabilities from current $(I_{it}, \widetilde{I}_{it})$ to next-period $(I, \widetilde{I})$ is independent from firm j. That is, $Pr(\boldsymbol{D}|j, \tau, I_{it} \cup \widetilde{I}_{it}) \equiv Pr(\boldsymbol{D}|\tau, I_{it} \cup \widetilde{I}_{it})$.

[64]This point is also true if workers have additional information about themselves. There can be self-selection of workers into firms with higher returns to publications based on workers' private information.

B. Data

Doctoral Dissertations. Appendix Table B2 displays the number of dissertations by year. For school×year cells with particularly low or missing data on ProQuest, I collected about 15,000 more Ph.D. profiles from school-specific sources, such as department websites or dissertation repositories. For example, the number of new dissertations from Carnegie Mellon University dropped from 100 to 30 in 2014. I then collected additional dissertations from its own open-access repository KiltHub. See a detailed breakdown of dissertations on ProQuest versus school-specific sources in Appendix Table B3. The total number of Ph.D. graduates in the sample by year, which stays around 3,000-3,300 per year from the top 60 schools since 2006 (Appendix Figure B4).

LinkedIn Profiles With the Recruiter Lite account, LinkedIn allowed me to view public profiles within my third degree of connections. To deal with this limitation, I actively connected with a random sample of Ph.D. graduates before the web scraping for each school. I connected with individuals who published at CS conferences, or research scientists at various companies. If an individual is on LinkedIn but falls outside my 3rd-degree connections, the search result would indicate "Out of Network". There were about 1,800 out-of-network profiles in total, out of fifty thousand queries that returned at least one profile on LinkedIn. I manually checked a random sample of out-of-network profiles and found that most of them had less than 100 connections on LinkedIn.

Publications Data. The main data source of research papers is Scopus, an abstract and citation databases of peer-reviewed literature launched by Elsevier in 2004. For each conference/journal × year, a query is submitted via Scopus Search API, and it returns a list of papers with information such asauthor(s), title, abstract, ISSN, DOI, number of citations, volume, issue, and publication date. Scopus also provides affiliations IDs at paper × author level. Another query is submitted for each affiliation ID via the Affiliation Search API, and returns the corresponding institution's name and location. I consider a paper by author i affiliated to j as her on-the-job research if (1) j can be matched with an employer of i on her LinkedIn profile; and (2) Author i is employed by j at the time of publication.

If a paper has multiple authors, I flag the paper if the majority of coauthors come from i's Ph.D. institution, which is likely to indicate a publication of her dissertation rather, especially if it happens within the first year after PhD. I also flag papers where coworkers come from a different industry employer, and remove papers that are matched with a worker's previous employer rather than her current one. For example, a person who moves from Yahoo to Microsoft might put Microsoft as her affiliation at the time of publication, but if her coauthors come from Yahoo, it is likely to indicate a work done at Yahoo rather than Microsoft. Typically this kind of papers would declare "This work was done when X was at ...".

To evaluate paper quality, I collected citations from Scopus, which covers both journal articles and conference papers. Citations from other conference papers are particularly important in computer science. For each paper, I counted the number of citations by year since publication, excluding self-citations by the authors themselves. Papers with a matched patent application receive more citations over time as shown in Figure 3. The citations on Scopus are mostly conference papers or journal articles.

Figure B1: Publication of Patent Applications that are Matched to CS Papers
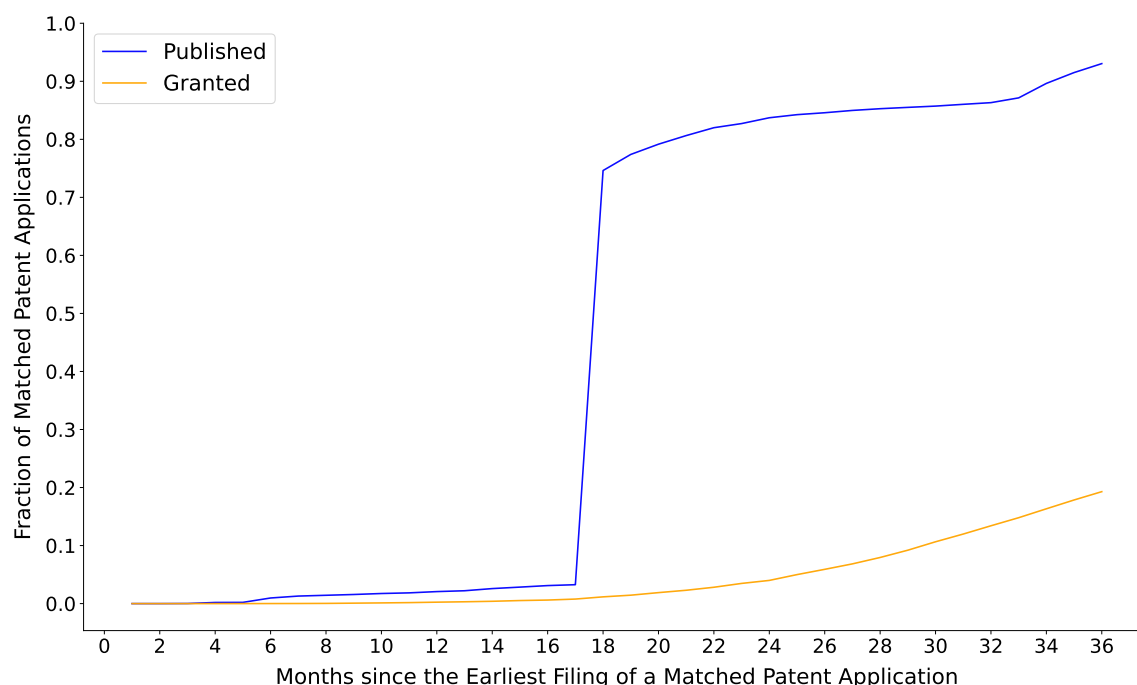
Notes: This figure shows the fraction of patent applications matched to a CS paper that have been published (blue) or granted (yellow) by month since the earliest patent filing date. The jump in the share published at 18 months since the initial filing is consistent with the 18-month rule in 35 U.S.C. 122 since the American Inventors Protection Act (AIPA 1999). About 20% of matched patent applications are disclosed later than 18 months. An audit study suggests that the non-compliance is driven by applicants who file a non-publication request at the time of the initial filing, as explained by Exception B of 35 U.S.C. 122 (b) in Table B3. Such applications will be published when the US patent office makes a final decision about whether a patent can be issued or the application should be rejected. Looking at three years since the earliest filing, more than 95% of matched patent applications have been published.

Figure B2: Job Postings for Research Scientists

(a) Amazon Science

BASIC QUALIFICATIONS

- Graduate degree (MS or PhD) in Computer Science, Electrical Engineering, Mathematics or Physics
- Minimum 3+ years of research experience or 2+ years of work experience developing and commercializing computer vision or deep learning
- 2+ years of experience implementing computer vision or deep learning algorithms in C++, C, Python or equivalent programming languages
- 2+ years of experience developing deep learning algorithms including but not limited to few-shot learning, zero-shot learning, foundational models, transfer learning.

PREFERRED QUALIFICATIONS

- Experience with conducting research in a corporate setting
- Excellent publication record in peer reviewed conferences and journals
- Proven expertise in conducting independent research and building computer vision systems.
- Experience working in the intersection of vision and language
- Proficient in C++ and Python, and familiar with non-linear optimization/filtering algorithms.

(b) Google Research

Minimum qualifications:

- PhD in Computer Science, related technical field or equivalent practical exp
- Experience in Natural Language Understanding, Computer Vision, Machine Optimization, Data Mining or Machine Intelligence (Artificial Intelligence).
- Programming experience in C, C++, Python.
- Contributions to research communities/efforts, including publishing papers NeurIPS, ICML, ACL, CVPR).

Preferred qualifications:

- Relevant work experience, including full time industry experience or as a re
- Strong publication record
- Ability to design and execute on research agenda.

Notes: This figure shows recent postings of research scientist jobs at Amazon and Google. Both ads explicitly indicate a graduate degree in computer science as a basic qualification for this type of jobs, and list "publication records" as preferred qualifications.

Figure B3: CS PhDs in NSF Surveys

(a) Num. New PhDs

New PhDs by Field (Survey of Earned Doctorates)

(b) % Employed by Business, Age 30-34
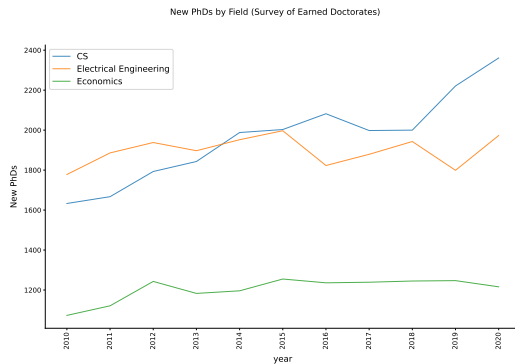
Fraction Employed by Business, Age Age 30-34
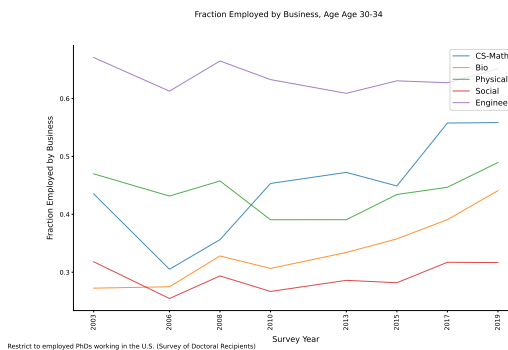
Notes: (a) displays the number of new PhDs in the Survey of Earned Doctorates by NSF. (b) come from the the Survey of Doctoral Recipients, restricted to Ph.D. recipients in the U.S. with nonmissing employer information between age 30-34.

Figure B4: Number of PhD Dissertations and Matched LinkedIn Profiles by Graduation Year
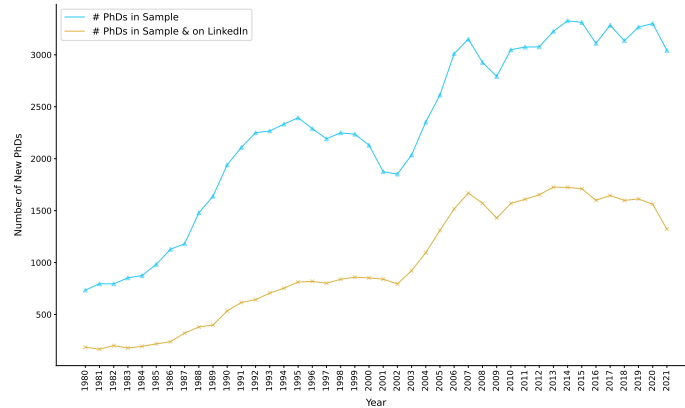
Notes: The blue line (top) shows the number of Ph.D. recipients in Computer Science or Electrical Engineering identified in ProQuest dissertation database or various school-specific sources (Appendix Table B2) by graduation year from 1980 to 2021. The yellow line plots the number of Ph.D.s who are matched with a public LinkedIn profile by full name, Ph.D. institution, year of graduation.

Figure B5: LinkedIn Platform
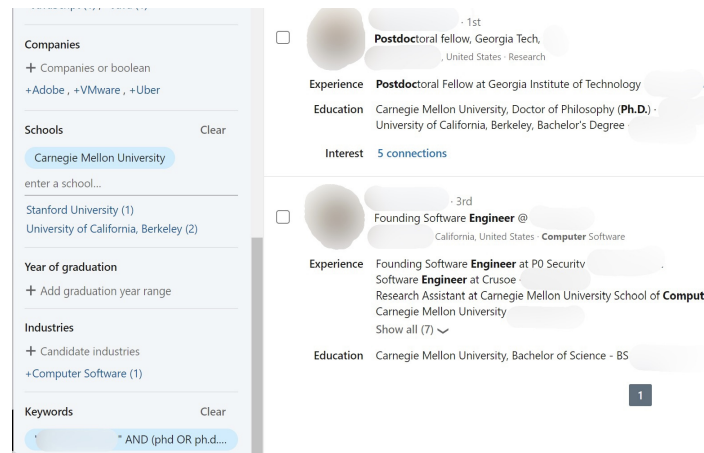
Notes: This figure shows the outputs of one query on the LinkedIn Recruiter Lite platform. The query includes the full name of a CS Ph.D. and keywords about a "Ph.D." degree and about CS such as "computer science" or "electrical engineering". The search is also restricted to CMU, where the person receives the Ph.D. degree. This query returns two profiles. The first profile returned perfectly matches the name and education info, whereas the second person has a very different name. If the fuzzy partial text match score between the actual full name and that on a LinkedIn profile falls below 50 (out of 100), the scraper would not collect that profile.

Figure B6: ROC Curve for Paper-Patent Matching by Threshold of Embedding Distance
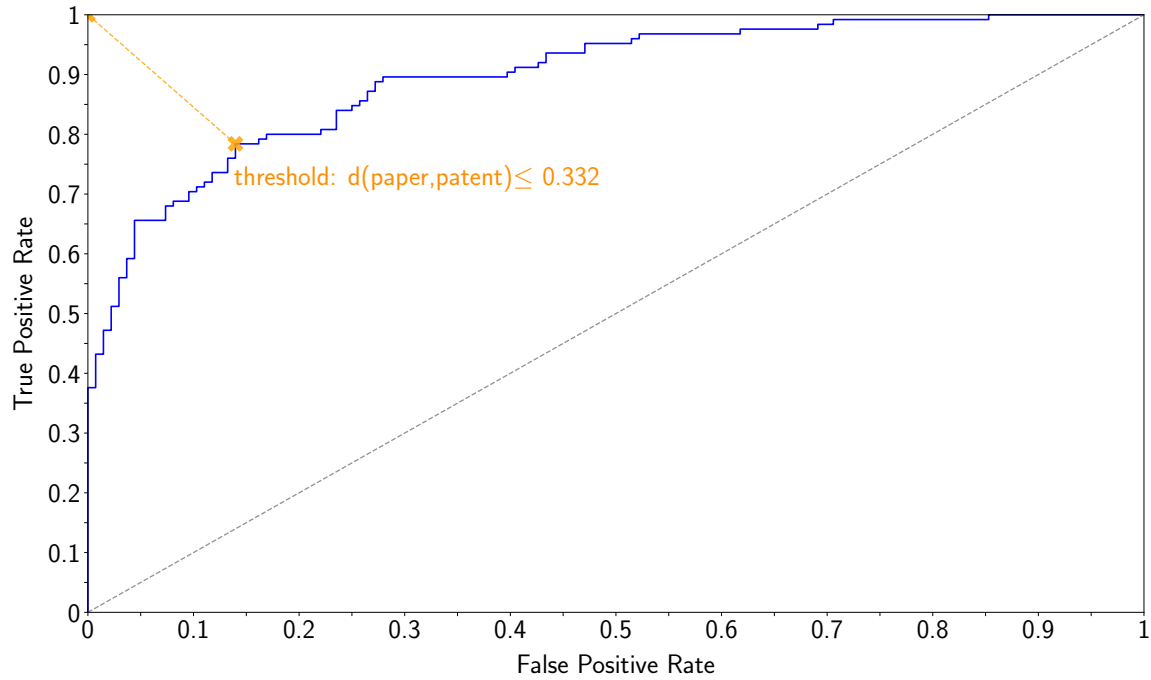
Notes: A paper and a patent application are defined as a match if they are produced by almost the same researchers at the same institution and discuss almost identical research findings from the same project. This figure shows the ROC curve of a predictor for paper-patent matches based on the distance between a paper's embedding and a patent application's embedding. A paper-patent is predicted as a match if the distance falls below a certain threshold. The performance of this classification model is evaluated on a random sample of 200 paper-patent pairs that satisfy the other three criteria (see Section 4.3.2). By reading the complete text of papers and patent applications rather than just titles and abstracts, I manually labeled the true matches. We then calculated the true positive rates (recall) and false positive rates of the predictor at each threshold, and selected 0.35 as the threshold that is relatively closer to the most desirable $(0, 1)$.

Table B1: Explanatory Power of PhD School + Cohort Fixed Effects

Economics		CS/EE	
Outcome	R^2	Outcome	R^2
Ln Citations Pre Tenure	0.275	Ln Citations in 5 Yrs	0.063
Num. Papers Pre Tenure	0.188	Num. Papers in 5 Yrs	0.055

Note: Economist CV data is provided by Sarsons (2017).

Table B2: Number of Profiles by Year

	PhD Graduates		LinkedIn Profiles		
Yr. PhD	ProQuest	Websites	Potential Profiles	Out-of-network	Matched Profiles
1980	595	140	254	11	185
1981	640	156	241	25	166
1982	639	156	272	23	200
1983	662	191	250	18	178
1984	702	173	285	25	193
1985	772	211	335	38	218
1986	920	208	384	45	238
1987	1002	179	432	26	321
1988	1393	85	559	40	380
1989	1571	68	610	61	399
1990	1873	68	717	50	535
1991	2040	69	832	58	616
1992	2162	88	859	65	643
1993	2179	88	923	61	706
1994	2244	89	981	59	753
1995	2303	91	1066	56	813
1996	2190	99	1097	79	819
1997	2100	92	1043	51	801
1998	2158	91	1116	59	839
1999	2151	85	1099	48	859
2000	2038	92	1104	51	853
2001	1778	97	1064	52	840
2002	1764	88	990	44	795
2003	1924	112	1138	43	922
2004	2194	159	1322	44	1095
2005	2462	152	1645	62	1310
2006	2779	232	1892	65	1516
2007	2900	251	2087	67	1669
2008	2726	201	1967	60	1571
2009	2499	293	1792	42	1429
2010	2508	541	1932	48	1570
2011	2500	575	1965	46	1609
2012	2523	554	2046	31	1653
2013	2426	801	2133	25	1726
2014	2388	940	2215	28	1724
2015	2274	1038	2213	44	1711
2016	2258	853	2084	27	1599
2017	2266	1019	2182	24	1646
2018	2197	939	2086	26	1598
2019	2107	1160	2118	37	1613
2020	2193	1108	2035	43	1561
2021	1971	1071	1823	38	1321

Table B3: Number of Profiles by School (ProQuest vs. School-specific Dissertation Database or Websites)

School	ProQuest Dissertations			School-specific Sources		
	# Dissertations	LinkedIn Profiles	Matched	# Dissertations	LinkedIn Profiles	Matched
austin	2028	990	845	1671	762	635
berkeley	3169	1949	1618	836	369	272
caltech	721	435	296	402	184	112
cmu	2357	1537	1259	2332	920	695
cornell	1738	962	685	481	203	125
git	2379	1426	1174	2300	1230	946
maryland	2421	1380	1143	895	233	169
michigan	2520	1403	1082	1052	331	244
mit	3726	2259	1684	769	353	251
nyu	478	272	200	147	58	48
oregon	412	196	144	233	157	76
princeton	1297	818	637	88	44	35
psu	1734	1012	807	181	91	65
purdue	2448	1387	825	202	87	77
rutgers	837	507	377	350	103	64
ucsb	1450	904	758	61	20	15
uiuc	3541	2070	1630	2359	776	451
umass	826	480	336	296	192	131
utah	714	418	296	48	20	12

Table B4: Patent Laws - Title 35, United States Code

Law	Content
35 U.S.C. 102	CONDITIONS FOR PATENTABILITY
(a)	NOVELTY; PRIOR ART.- A person shall be entitled to a patent unless—
(A)	the claimed invention was patented, **described in a printed publication**, ..., or otherwise available to the public **before the effective filing date** of the claimed invention
(b)	EXCEPTIONS: **(1) A disclosure made 1 year or less before the effective filing date** of a claimed invention shall not be prior art to the claimed invention under subsection (a)(1) if—
(A)	the disclosure was **made by the inventor or joint inventor or by another** who obtained the subject matter disclosed **directly or indirectly from the inventor or a joint inventor**; or
(B)	the subject matter disclosed had, before such disclosure, been publicly **disclosed by the inventor or a joint inventor** or another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor.
35 U.S.C. 122	CONFIDENTIAL STATUS OF APPLICATIONS; PUBLICATION OF PATENT APPLICATIONS
(a)	CONFIDENTIALITY.— Except as provided in subsection (b), applications for patents shall be kept in confidence by the Patent and Trademark Office and no information concerning the same given without authority of the applicant or owner unless necessary to carry out the provisions of an Act of Congress or in such special circumstances as may be determined by the Director.
(b)	PUBLICATION.-
(1)	IN GENERAL.— (A) Subject to paragraph (2), each application for a patent shall be published, ..., promptly after the expiration of a period of **18 months from the earliest filing date** for which a benefit is sought under this title.
(2)	EXCEPTIONS.— (A) (i) no longer pending; (ii) subject to a secrecy order under section 181 ; (iii) a **provisional application** filed under section 111(b); or (iv) an application for a design patent...
(2)	EXCEPTIONS.- (B) If an applicant makes a request upon filing, certifying that the invention disclosed in the application has not and will not be the subject of an application filed in another country...

Notes: Detailed discussions of title 35 U.S.C. can be found on the USPTO websites: U.S.C. 102 pre-AIA, U.S.C. 102 AIA, U.S.C. 122. Notably, the America Invents Act in 2011 switched the U.S. patent system from a "first to invent" to a "first to file" system. But the 12-month grace period in filing a patent application for inventors' own publications (35 U.S.C. 102), and the 18-month publication rule (35 U.S.C. 122) have not changed since the American Inventors Protection Act (AIPA 1999).

Table B5: Descriptive Statistics: Matched Computer Scientists

	Full Sample		Balanced sample	
	Mean	SD	Mean	SD
Gender from Name or Picture				
Female	0.118	0.323	0.123	0.329
Male	0.725	0.446	0.708	0.455
Education				
Year of Ph.D.	2007	9.853	2011	3.689
Ph.D. in CS (∋ EECS)	0.531	0.499	0.522	0.500
Ph.D. in EE	0.469	0.499	0.478	0.500
If bachelor information is available:				
Bachelor in the U.S.	0.446	0.497	0.386	0.487
Bachelor from Top 60 CS in the U.S.	0.288	0.453	0.249	0.432
Research Outputs Post Ph.D.				
Num. Papers	2.506	9.452	2.491	8.767
Num. Paper-Patent Matches	0.219	1.444	0.231	1.413
Num. Patent Applications Not Matched to a Paper	1.672	3.142	1.375	2.275
Any Paper	0.282	0.450	0.297	0.457
Any Paper-Patent Match	0.067	0.250	0.074	0.261
Any Patent Application Not Matched to a Paper	0.426	0.494	0.448	0.497
Employment Post Ph.D.				
Num. Yrs with Full-time Employment	13.498	6.910	11.530	3.692
Num. Tenure-track Employers	0.300	0.617	0.259	0.574
Num. Postdoc Employers	0.154	0.398	0.205	0.454
Num. Top Firms	0.295	0.541	0.373	0.598
Num. Nontop Firms	1.866	1.664	1.612	1.310
Ever on the Tenure track	0.231	0.421	0.198	0.398
Ever a Postdoc	0.141	0.348	0.185	0.388
Ever at Top Firms	0.256	0.436	0.316	0.465
Ever at Nontop Firms	0.795	0.404	0.800	0.400
Observations	40,219		18,860	

Notes: This table summarizes the sample of matched Ph.D.'s with non-missing full-time employment records on LinkedIn (Section 3.2). The full sample (first two columns) includes matched CS/EE Ph.D.'s from top 60 CS schools who graduated between 1980 and 2021, and have at least one full-time job with one employer self-reported on LinkedIn. We use the full sample throughout Section 4. The balanced (sub)sample restricts to those who graduated between 2005 and 2018 and have 5 years of non-missing job history since Ph.D. on LinkedIn. We use this subsample to estimate the 5-period structural model in Section 5.
• Gender is classified based on either first name or profile picture (available for 78% of the sample). 15% remains missing, due to either a missing picture or gender-neutral or foreign names that cannot be classified based on the U.S. Census.
• A paper-patent match is established according to 3.3.2.

Table B6: Descriptive Statistics: Person-Year Panel

$j(i,t) \in$	Nontop Firms		Top Firms		Academia	
	Mean	SD	Mean	SD	Mean	SD
Experience (Years since Ph.D.)	11.678	8.569	9.209	7.322	11.587	9.052
Experience in Academia	1.171	3.236	0.675	2.222	9.771	8.173
Tenure	5.007	5.449	4.981	5.352	7.575	7.672
Current Position						
Tenure-track	0.000	0.009	0.000	0.000	0.728	0.445
Postdoc	0.000	0.000	0.000	0.000	0.104	0.305
Research Scientist	0.119	0.324	0.149	0.356	0.036	0.186
Engineer	0.453	0.498	0.604	0.489	0.036	0.187
Manager	0.153	0.360	0.195	0.396	0.016	0.127
Senior Position	0.496	0.500	0.391	0.488	0.053	0.224
Any Promotion	0.062	0.242	0.064	0.245	0.060	0.238
Research Outputs						
Any Paper	0.023	0.151	0.113	0.317	0.185	0.388
Any Paper-Patent Match	0.006	0.075	0.033	0.180	0.013	0.111
Any Patent App Not Matched to a Paper	0.126	0.332	0.203	0.402	0.047	0.212
Movements between Employers $j(i,t)$ vs. $j(i,t+1)$						
New Employer Next Year	0.118	0.323	0.065	0.247	0.074	0.262
Employed by Top Firms Next Year	0.016	0.124	0.949	0.221	0.006	0.079
Observations	331,451		68,230		143,197	

Notes: This table summarizes the person×year level panel for matched Ph.D.'s. The first two columns display the means across person×year observations for those currently employed by a firm outside the top tier in the industry, denoted as $j(i,t) \in$ non-top. The second set restricts to those working at top firms, and the third set to those working in academia (including postdocs, tenure-track jobs or other roles). We put all postdocs and faculty in the third group. There are 530 person×year observations (226 individuals) where a person works as a postdoc or visiting scholar in one of the top firms.

Table B7: Examples of CS Papers and Matched Patent Applications

Firm	Team Over-lap	Text Distance	Papers		Matched Patent Applications		
			Title	M/Yr	Title	Filing M/Yr	Published M/Yr
Microsoft	100%	0.247	FROID OPTIMIZATION OF IMPERATIVE PROGRAMS IN A RELATIONAL DATABASE	12/2017	METHOD FOR OPTIMIZATION OF IMPERATIVE CODE EXECUTING INSIDE A RELATIONAL DATABASE ENGINE	05/2017	11/2018
Adobe	80%	0.273	FORECASTING HUMAN DYNAMICS FROM STATIC IMAGES	07/2017	FORECASTING MULTIPLE POSES BASED ON A GRAPHICAL IMAGE	04/2017	10/2018
Google	70%	0.146	VARIABLE RATE IMAGE COMPRESSION WITH RECURRENT NEURAL NETWORKS	05/2016	IMAGE COMPRESSION WITH RECURRENT NEURAL NETWORKS	02/2016	01/2019
Yahoo	100%	0.233	UNBIASED ONLINE ACTIVE LEARNING IN DATA STREAMS	08/2011	ONLINE ACTIVE LEARNING IN USER-GENERATED CONTENT STREAMS	10/2011	05/2013
IBM	100%	0.121	A TAG BASED APPROACH FOR THE DESIGN AND COMPOSITION OF INFORMATION PROCESSING APPLICATIONS	09/2008	FACETED, TAG-BASED APPROACH FOR THE DESIGN AND COMPOSITION OF COMPONENTS AND APPLICATIONS IN COMPONENT-BASED SYSTEMS	10/2008	04/2010

Notes: This table presents examples of CS papers and matched patent applications. "Firm" refers to the common affiliation of authors, which is matched to the assignee of the matched patent. "Team Overlap" is defined as the fraction of inventors on a patent application who are matched with authors on the paper. Research assistants or interns may be authors on a paper but excluded from inventors on a patent application. "Text distance" is measured by the distance between the embedded vector for a paper's title and abstract, and that of a patent's. The word embedding was done via OpenAI's Ada V2 model. The timestamp "M/Yr" for a paper is the month/yr when it is published at a conference. "Filing M/Yr" for a patent application is based on the earliest filing or priority date, and in "Published M/Yr" a patent application becomes public for the first time.

C. Reduced-Form Tests for Employer Learning (Section 4)

Figure C1: Heterogeneity in Mobility Responses by Experience since PhD

(a) Any Move between Firms

(b) Employment by Top Firms Next Year
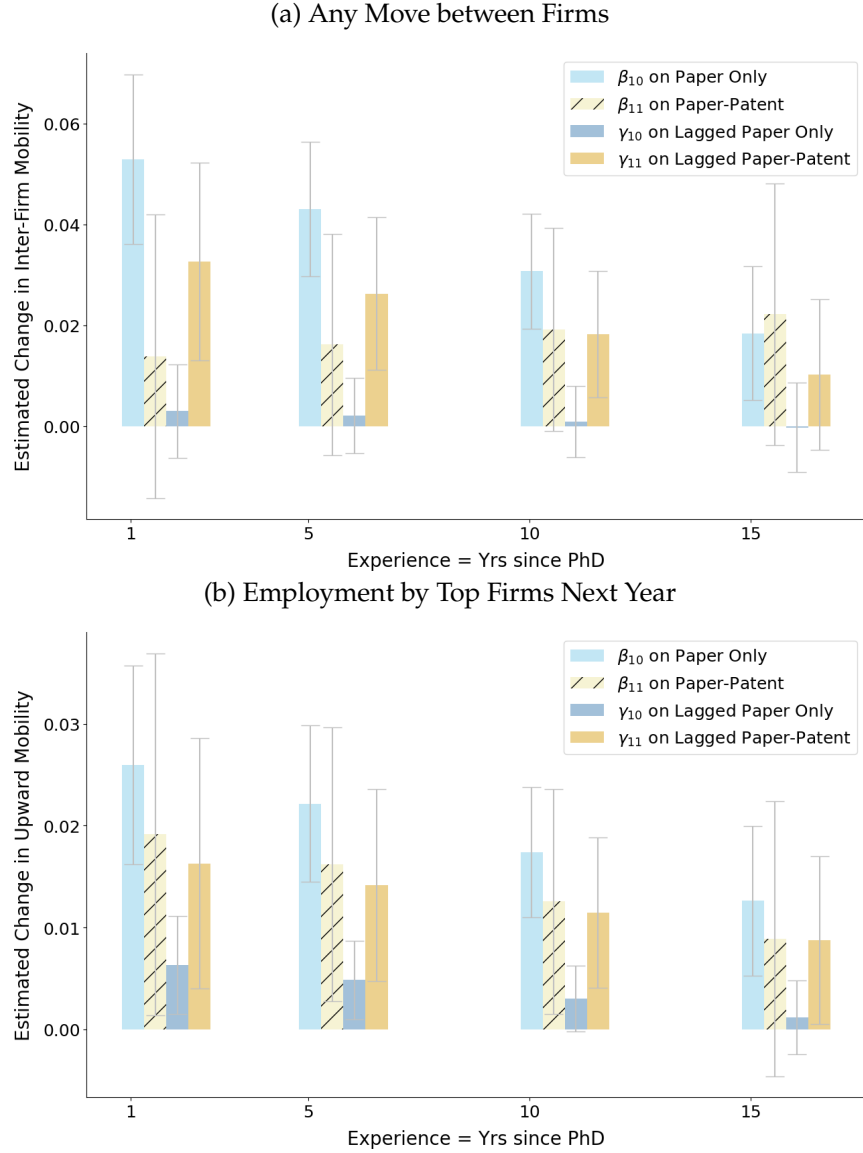
Notes: I add interactions between $D_{it}(P)$, $D_{it}(PQ)$, Lagged-$D_{it}(P)$, Lagged-$D_{it}(PQ)$ and years of experience since PhD to the regression (4.1). The barplot above shows the estimated $\hat{\beta}_k$ on $D_{it}(k)$ and $\hat{\gamma}_k$ on Lagged-$D_{it}(k)$ for $k = P, PQ$, respectively at each experience level.

Table C1: Job Mobility on Papers & Matched Patents (Poisson Regressions)

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
CS Papers at t : $D_{it}(P)$ vs. $D_{it}(PQ)$						
Paper only	0.2626	-0.0227	0.0992	0.5395	0.0034	0.3048
	(0.0382)	(0.0617)	(0.0304)	(0.0800)	(0.0038)	(0.0985)
Paper+Matched Patent	0.1495	0.0145	0.0128	0.3251	0.0021	0.3290
	(0.0640)	(0.0810)	(0.1016)	(0.1234)	(0.0058)	(0.2274)
CS Papers in [t − 3, t − 1]: Lagged-$D_{it}(P)$ vs. Lagged-$D_{it}(PQ)$						
Paper only	0.0083	0.0134	0.1153	0.1052	-0.0003	0.6870
	(0.0274)	(0.0463)	(0.0270)	(0.0550)	(0.0030)	(0.0957)
Paper+Matched Patent	0.1393	0.0910	0.0598	0.2593	0.0003	0.7869
	(0.0426)	(0.0661)	(0.0714)	(0.0915)	(0.0050)	(0.1818)
Patents unrelated to CS Papers						
$D_{it}(Q)$	-0.1114	-0.0712	-0.0990	-0.0175	0.0089	0.1389
	(0.0189)	(0.0415)	(0.0500)	(0.0473)	(0.0027)	(0.1120)
Lagged-$D_{it}(Q)$	0.0417	-0.0189	0.0749	0.1194	0.0035	0.0360
	(0.0148)	(0.0363)	(0.0345)	(0.0401)	(0.0022)	(0.1081)
Mean	.1588418	.0656451	.1209954	.0469412	.9485002	.0304762
N	161K	66K	75K	86K	66K	27K
Pseudo R^2	.1377074	.0382099	.1894513	.1777506	.0003756	.2066823

Notes: This table presents Poisson regressions of the mobility outcomes (indicators) on the same controls and fixed effects as specified in (4.3). The coefficients on $D_{it}(k)$ or Lagged-$D_{it}(k)$ for $k = P, PQ, Q$ represent proportional increase in job mobility among workers with output k relative to coworkers group without an innovation output. Observations that are separated by a fixed effect are dropped from the estimation sample of a Poisson regression. For example, if the mean of the dependent variable is 0 at a firm-yr (j, t), all observations within that (j, t) would be dropped in Poisson regression above but not in OLS (Table 3).

Table C2: Effects of Papers & Matched Patents on Job Mobility (Person Fixed Effect)

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
CS Papers at t : $D_{it}(P)$ vs. $D_{it}(PQ)$						
Paper only	0.0325	-0.0040	0.0065	0.0113	0.0055	0.0011
	(0.0063)	(0.0045)	(0.0029)	(0.0034)	(0.0039)	(0.0010)
Paper+Matched Patent	0.0309	0.0045	0.0025	0.0127	0.0026	0.0014
	(0.0115)	(0.0067)	(0.0070)	(0.0060)	(0.0056)	(0.0026)
CS Papers in $[t-3, t-1]$: Lagged-$D_{it}(P)$ vs. Lagged-$D_{it}(PQ)$						
Paper only	0.0066	-0.0022	0.0082	-0.0007	0.0012	0.0038
	(0.0043)	(0.0038)	(0.0027)	(0.0023)	(0.0034)	(0.0011)
Paper+Matched Patent	0.0306	0.0110	0.0048	0.0080	0.0041	0.0040
	(0.0080)	(0.0065)	(0.0057)	(0.0043)	(0.0058)	(0.0022)
Patents unrelated to CS Papers						
$D_{it}(Q)$	0.0044	0.0089	-0.0007	0.0015	-0.0030	0.0017
	(0.0025)	(0.0031)	(0.0043)	(0.0011)	(0.0028)	(0.0015)
Lagged-$D_{it}(Q)$	0.0183	0.0140	0.0036	0.0033	-0.0100	-0.0020
	(0.0024)	(0.0030)	(0.0033)	(0.0011)	(0.0027)	(0.0010)
Mean	.1105	.0624	.0683	.0167	.9521	.0058
N	222K	65K	121K	222K	65K	121K
Adj. R^2	.1993	.0969	.1883	.1404	.0969	.1718

Notes: This table presents regression estimates of equation 4.1 with person fixed effects. See the notes under Table 3 for details on other controls.

Table C3: Additional Mobility Outcomes - Wage Growth and Academic Employment

	Move to a Higher-Wage Firm		Higher-Wage Position		Move to Academia		
	(1) Nontop	(2) Top	(3) Nontop	(4) Top	(5) Nontop	(6) Top	(7) Academia
CS Papers at t : $D_{it}(P)$ vs. $D_{it}(PQ)$							
Paper only	0.0280	-0.0005	0.0313	0.0060	0.0139	0.0074	0.0185
	(0.0056)	(0.0035)	(0.0078)	(0.0039)	(0.0026)	(0.0019)	(0.0019)
Paper+Matched Patent	0.0209	-0.0014	0.0115	0.0130	0.0056	0.0091	0.0122
	(0.0093)	(0.0056)	(0.0117)	(0.0076)	(0.0041)	(0.0031)	(0.0057)
CS Papers in $[t-3, t-1]$: Lagged-$D_{it}(P)$ vs. Lagged-$D_{it}(PQ)$							
Paper only	0.0017	-0.0018	-0.0021	-0.0029	0.0051	0.0028	0.0107
	(0.0032)	(0.0024)	(0.0041)	(0.0024)	(0.0013)	(0.0012)	(0.0018)
Paper+Matched Patent	0.0132	0.0072	0.0174	0.0014	0.0077	-0.0008	0.0179
	(0.0059)	(0.0044)	(0.0080)	(0.0046)	(0.0027)	(0.0020)	(0.0038)
Mean	.0594277	.039501	.0428258	.026157	.0099243	.0058593	.9498891
N	131K	59K	52K	45K	220K	66K	122K
Adjusted R^2	.087463	.0185282	.0625471	.0178933	.0934459	.0076865	.0478011

Notes: This table presents estimates of 4.1 for changes in job titles as reported on LinkedIn. The first three columns show the regression of any promotion on innovation outputs $D_{it}(k)$, Lagged-$D_{it}(k)$ for $k = P, PQ$, which is estimated on workers who are not in senior roles yet (e.g., not a "senior software engineer"). In academia, a promotion is coded as assistant professors getting tenured. Columns (4)-(9) are estimated for workers in the industry. (4)-(5) present the regressions of becoming a research scientist on innovation outputs, estimated on employees who are not research scientists at nontop firms , and at top firms, respectively. Likewise, becoming an engineer or manager is estimated on workers who are not an engineer or manager yet.

Table C4: Additional Mobility Outcomes - Promotion | Stayers

	Promotion			New Scientist		New Engineer		New Manager	
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Nontop	(7) Top	(8) Nontop	(9) Top
CS Papers at t : $D_{it}(P)$ vs. $D_{it}(PQ)$									
Paper only	0.0413	0.0370	0.0470	0.0090	-0.0047	0.0058	-0.0035	0.0078	0.0082
	(0.0065)	(0.0056)	(0.0031)	(0.0050)	(0.0032)	(0.0034)	(0.0026)	(0.0029)	(0.0028)
Paper+Matched Patent	0.0324	0.0120	0.0478	0.0194	0.0042	-0.0033	-0.0038	0.0136	0.0038
	(0.0124)	(0.0070)	(0.0101)	(0.0105)	(0.0044)	(0.0038)	(0.0032)	(0.0054)	(0.0046)
CS Papers in $[t-3, t-1]$: Lagged-$D_{it}(P)$ vs. Lagged-$D_{it}(PQ)$									
Paper only	0.0081	0.0034	0.0119	0.0061	0.0026	-0.0052	-0.0007	-0.0017	-0.0021
	(0.0035)	(0.0034)	(0.0025)	(0.0032)	(0.0034)	(0.0019)	(0.0022)	(0.0016)	(0.0019)
Paper+Matched Patent	0.0278	0.0077	0.0200	0.0120	-0.0038	0.0004	-0.0035	0.0013	0.0056
	(0.0086)	(0.0058)	(0.0069)	(0.0077)	(0.0037)	(0.0036)	(0.0026)	(0.0033)	(0.0035)
N	87K	37K	65K	172K	53K	88K	24K	160K	49K
Adjusted R^2	.040206	.0220636	.0461156	.1642801	.0366538	-.0251746	.0032157	.0111164	.0077875

Notes: This table presents the same set of regressions of promotions or position changes on innovation outputs as in Table C4, but are estimated on stayers who are not moving to a new firm the next year.

D. Additional Results from Structural Analysis

Appendix Figure D1: Posterior versus Prior by PhD Ranking
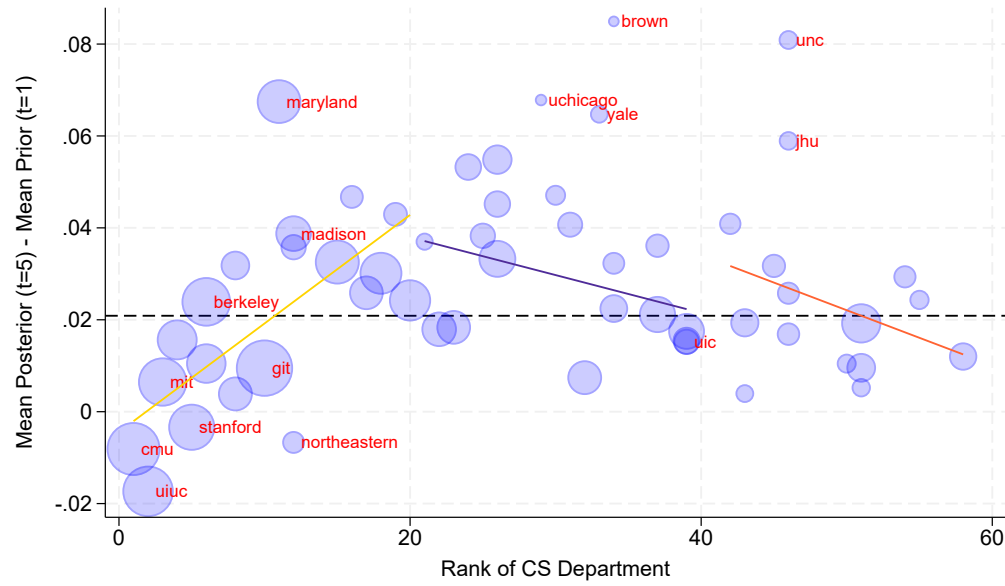
Appendix Figure D2: Allocation to Innovation Task against Employer Belief
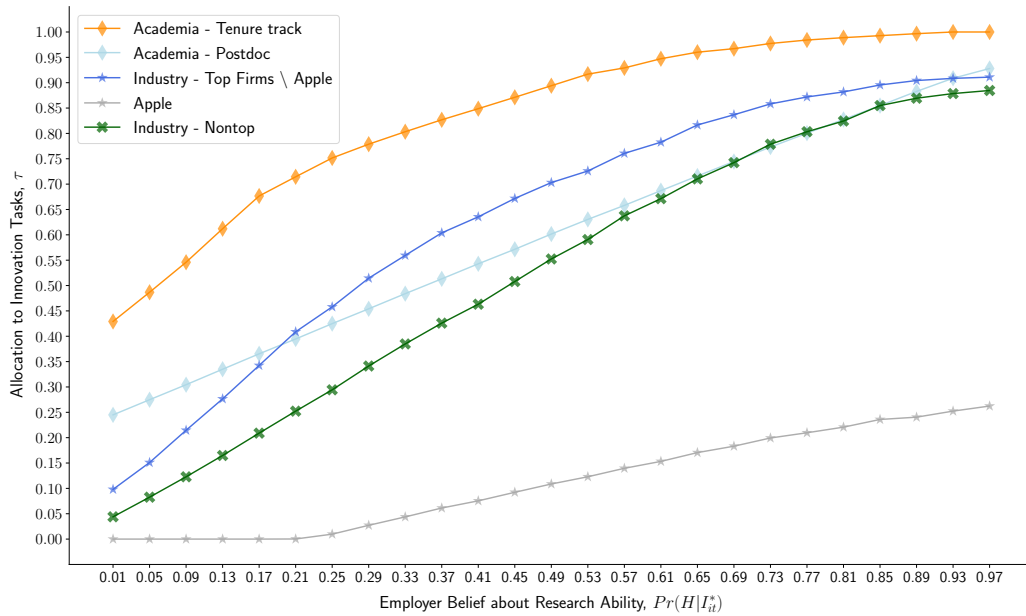
Appendix Figure D3: Change in Publication Rate in the Absence of Employer Learning

Counterfactual Publication Rate (rel. to benchmark)

Table D1: Model Parameters

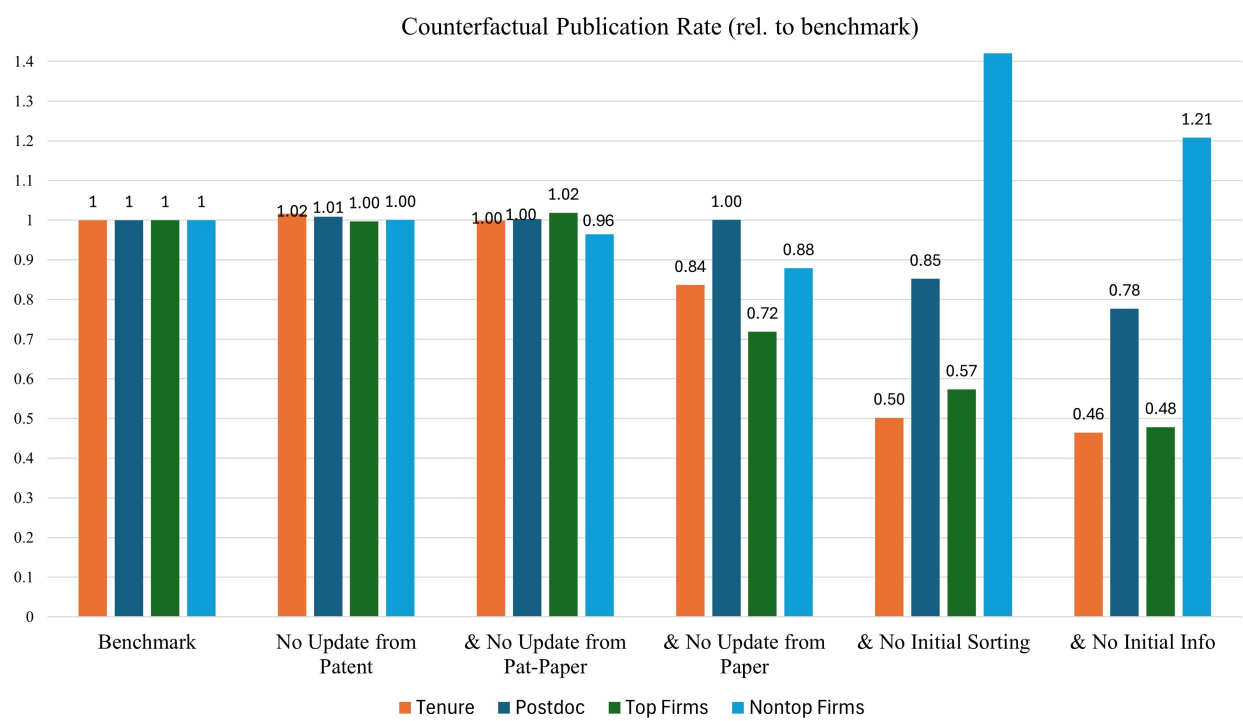| Parameter	Description	Calibration	Maximum-Likelihood Estimate
I. Common Prior			
δ	Logit Coefficient on $X(I_{i1})$ in (5.1)		$(-0.24, 0.009, 3.02, -2)$ on phd rank and pub before PhD $(-0.49, -0.98, -1.50, -2)$ on G_{i1}
II. Labor Supply - Preferences for Employers			
b	utility weight on log wage (2.6)		0.63
ρ_G	$1-$ corr. of ϵ_{itj} for $j \in$ nest G	$\rho_2 = 1$ for postdoc	$(0.78, 0.45, 0.88)$ at $G \neq 2$
$(\eta_{1,G}, \eta_{2,G})$	preference for market G: $\eta_{1,G}\pi + \eta_{2,G}\pi^2$	$(0.5, 1)$ at $G = 1$ $(0, 0)$ at $G = 2$	$(0.48, 0.49)$ at $G = 3$ $(-0.24, -0.49)$ at $G = 4$
$(\lambda_{0,G}, \lambda_{1,G})$	prob. of workers re-entering the labor market (2.5) $\lambda_G(\pi) = \lambda_{0,G} \times (1 + \lambda_{1,G} \times \pi)$, at $t > 1$	$(0.40, -0.50)$ at $G = 2$	$(0.04, -0.5)$ at $G = 1$ $(0.08, 0.10)$ at $G = 3$, $(0.13, 0.99)$ at $G = 4$
$(\Lambda_{AJ}, \Lambda_{JA})$	prob. academia is open to workers from industry, and vice versa.		$(0.24, 0.32)$
III. Firm Productivity			
$\overline{\phi}_j$	Baseline productivity in routine tasks of 16 employers	$\bar\phi_1, \bar\phi_2, \bar\phi_3, \bar\phi_5, \bar\phi_{11}$	Table D2
$\phi_j(P)$	j's proportional return to paper		Table D2
$\phi_j(Q)$	j's proportional return to patent	j-fixed effect in patenting	Table D2
$\phi_j(PQ)$	j's proportional return to paper-patent	$\phi_j(PQ) = 2.25\,\phi_j(P) + 0.25\,\phi_j(Q)$	
ζ	cost of innovation: $c(\pi, \tau) = \frac{\zeta}{2}\tau^2$	$\zeta = 6$	
IV. Worker Productivity			
p_H, p_L	prob. of a H-ability producing a paper ($y = 1$)		$(0.81, 0.19)$
$\widetilde{p}_H, \widetilde{p}L$	prob. of a L-ability producing a paper ($y = 1$)		$(0.42, 0.18)$
q_H, q_L	prob. of a H-ability producing a paper with a matched patent ($\widetilde{y} = 1$)		$(0.69, 0.51)$
Others			
β	exponential discount factor	0.90	

Notes: There are 56 free parameters that are estimated by maximizing the joint likelihood of job movements and innovation outputs (5.2), using the limited-memory BFGS optimization algorithm (Liu and Nocedal 1989).

Table D2: Firm Level: Estimated Productivity, Size and Wage Returns

			Returns to Innovation	
j	Description	Baseline $\overline{\phi}_j$	Paper $\phi_j(P)$	Patent $\phi_j(Q)$
Nest 1. Academia - Tenure Track				
0	Nontop Schools	0.298	9.699	1.605
1	Top 25 CS	0.011	16.001	1.840
Nest 2. Academia - Postdoc				
2	Postdoc at Nontop Schools	0.015	8.232	1.940
3	Postdoc at Top 25 CS	0.008	9.510	1.921
Nest 3. Industry - Top Firms				
4	IBM	0.005	9.791	10.659
5	Microsoft	0.022	7.306	5.142
6	Amazon	0.019	3.646	4.462
7	Facebook (Meta)	0.021	5.109	3.863
8	Apple	0.018	1.738	5.677
9	Google (Alphabet)	0.060	5.052	3.948
Nest 4. Industry - Nontop Firms (Grouped by Patenting FE)				
10	Above 90th Percentile	0.087	5.856	8.504
11	80th-90th	0.214	5.452	4.393
12	70th-80th	0.088	4.730	2.429
13	50th-70th	0.171	4.488	1.634
14	25th-50th	0.157	5.273	0.927
15	<25th Percentile	0.122	4.575	0.020

Notes: This table displays the maximum-likelihood estimates of the baseline productivity, $\overline{\phi}_j$, and their returns to publications, $\phi_j(P)$. The productivity in patenting, $\phi_j(Q)$, is calibrated based on the estimated firm fixed effect in filing patent applications unrelated to papers. I further calibrate the return to a paper with a matched patent as $\phi_j(PQ) = 2.25 \times \phi_j(P) + 0.25 \times \phi_j(Q)$.. In academia, "Top CS" includes the top 25 CS departments ranked by CSRankings: CMU, Berkeley, Stanford, MIT, Georgia Tech, Cornell, USC, UIUC, Princeton, Washington State, UCLA, UCSD, UMass - Amherst, UMich, Purdue, Maryland, Northeastern, Madison, Columbia, UT-Austin, UPenn, NYU, UC-Irvine, UC-Santa Barbara, UChicago, Stony Brook. Nontop firms in the industry are grouped according the the calibrated $\phi_j(Q)$.

Table D3: Descriptive Statistics: Person-Year Panel

$j(i, t) \in$	Nontop Firms		Top Firms		Academia	
	Mean	SD	Mean	SD	Mean	SD
Experience (Years since Ph.D.)	3.020	1.410	3.157	1.401	2.853	1.419
Experience in Academia	0.295	0.820	0.195	0.648	2.703	1.402
Tenure	2.020	1.495	2.203	1.533	1.898	1.432
Current Position						
Tenure-track	0.000	0.014	0.000	0.000	0.528	0.499
Postdoc	0.000	0.000	0.000	0.000	0.295	0.456
Research Scientist	0.170	0.376	0.167	0.373	0.047	0.212
Engineer	0.567	0.495	0.665	0.472	0.040	0.196
Manager	0.120	0.325	0.129	0.335	0.010	0.101
Senior Position	0.461	0.498	0.341	0.474	0.039	0.192
Any Promotion	0.097	0.296	0.089	0.284	0.057	0.231
Research Outputs						
Any Paper	0.042	0.200	0.128	0.334	0.206	0.404
Any Paper-Patent Match	0.011	0.104	0.041	0.198	0.013	0.115
Any Patent App Not Matched to a Paper	0.162	0.368	0.220	0.414	0.054	0.226
Movements between Employers $j(i, t)$ vs. $j(i, t+1)$						
New Employer Next Year	0.156	0.363	0.080	0.271	0.165	0.371
Employed by Top Firms Next Year	0.030	0.171	0.942	0.233	0.017	0.130
Observations	53,839		16,081		24,380	

Notes: This table summarizes the 5-yr balanced estimation sample at person×year level. We restrict to 18,860 workers who graduated between 2005 and 2018 and have full-time non-missing employment history for the first five years post PhD. See the notes under Table B5 and Table B6 for additional details on the variables.

Between-Within Decomposition

$$E[Pub_{it}|S \cup \{m\}] - E[Pub_{it}|S] \propto \sum_{\alpha} p_{\alpha} \times \underbrace{\left(\sum_{in \in \{0,1\}} \sum_{t,j,\pi} \triangle n_{tj,\alpha}^{(in)}(\pi) \times \tau_{tj}^{(in)}(\pi|S \cup \{m\}) \right)}_{\text{between firm}}$$

$$+ \sum_{\alpha} p_{\alpha} \times \underbrace{\left(\sum_{in \in \{0,1\}} \sum_{t,j,\pi} n_{tj,\alpha}^{(in)}(\pi|S) \times \triangle \tau_{tj}^{(in)}(\pi) \right)}_{\text{within firm}}$$

in which $\triangle n_{j,\alpha}^{(in)}(\pi) = n_{tj,\alpha}^{(in)}(\pi|S \cup \{m\}) - n_{tj,\alpha}^{(in)}(\pi|S)$ at state π

Table D4: Between-Within Decomposition

S	Publication Rate	% Change
Benchmark $\mathbb{M}$	0.0923	0%
$\mathbb{M}\backslash$ Employer Learning	0.0777	-15.8%
Hold $\tau(\cdot|\mathbb{M})$ *fixed:*		
$\mathbb{M}\backslash$ Employer Learning	0.0879	-4.8%